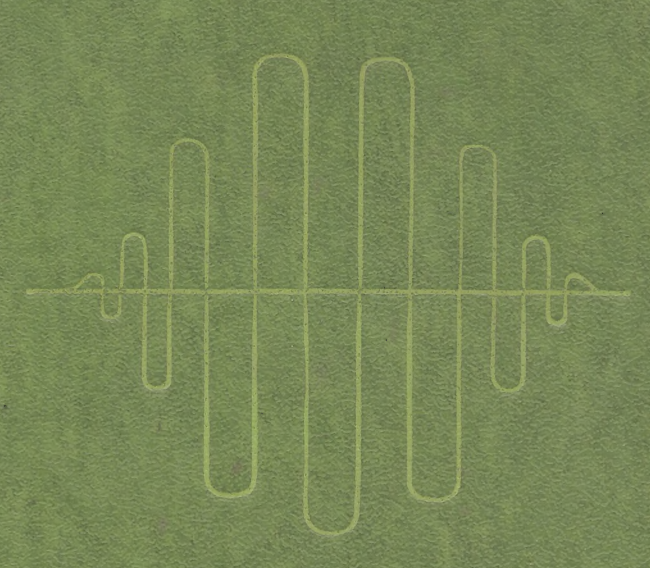
621.391
К89

И. В. КУЗЬМИН
В. А. КЕДРУС

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КОДИРОВАНИЯ



И. В. КУЗЬМИН, В. А. КЕДРУС

Стр. 263-268 - разорваны

АТу

065996

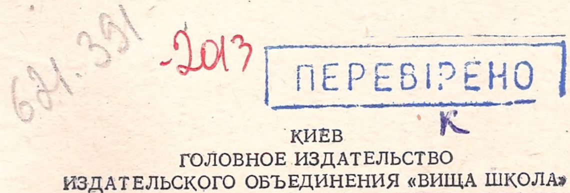
621.391
К89

И. В. КУЗЬМИН
В. А. КЕДРУС

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КОДИРОВАНИЯ

Допущено Министерством высшего и среднего специального образования УССР в качестве учебника для студентов вузов, обучающихся по специальностям «Автоматика и телемеханика» и «Прикладная математика»

065996



КИЕВ
ГОЛОВНОЕ ИЗДАТЕЛЬСТВО
ИЗДАТЕЛЬСКОГО ОБЪЕДИНЕНИЯ «ВИЩА ШКОЛА»
1977

6Ф0.1
К89

УДК 621.391.1

Основы теории информации и кодирования. Кузьмин И. В., Кедрус В. А. Киев, издательское объединение «Вища школа», 1977, 280 с.

В учебнике рассматриваются основы теории информации и кодирования, включающие общие понятия и определения информации и информационных систем, математическое представление детерминированных и случайных сигналов, информационные характеристики сигналов, эффективность и помехоустойчивость кодирования и передачи информации, теорию статистических решений и информационную оценку автоматизированных систем контроля и управления.

Учебник предназначен для студентов вузов специальностей «Автоматика и телемеханика» и «Прикладная математика». Может быть полезен широкому кругу специалистов, исследующих и разрабатывающих информационные системы.

Табл. 23. Ил. 144. Список лит.: 84 назв.

Редакция литературы по кибернетике, электронике и энергетике

Зав. редакцией *А. В. Дьячков*

К $\frac{30501-178}{М211(04)-77}$ 64—77

© Издательское объединение «Вища школа», 1977.

ПРЕДИСЛОВИЕ

В материалах XXV съезда КПСС отмечается, что первоочередной задачей остается ускорение научно-технического прогресса, так как только на основе ускоренного развития науки и техники может быть построено коммунистическое общество. Ускорение научно-технического прогресса обеспечивается непрерывным ростом объема научно-технической информации, которая удваивается через каждые 8—10 лет. В связи с этим ставится задача на разработку и внедрение на базе ЭВМ автоматизированных информационных систем, систем управления предприятиями, отраслями народного хозяйства и страной в целом.

В условиях, когда объем необходимых для инженера знаний быстро увеличивается, возрастает потребность в учебной литературе, которая должна помочь в процессе изучения основ научных теорий и обеспечить в дальнейшем на практике самостоятельное освоение новой информации.

Наиболее общей теорией сложных автоматизированных информационных систем, к которым относятся также автоматизированные системы управления, является теория информации.

Учебник написан для изучения курсов «Теоретические основы информационной техники» по специальности «Электронно-вычислительные машины», «Теория информации и кодирования» по специальности «Прикладная математика», «Теория информации» по специальности «Автоматизированные системы управления», «Теория передачи и преобразования информации» по специальности «Автоматика и телемеханика».

В учебнике девять глав. В первой главе дается понятие об информационных системах, основой функционирования которых являются процессы передачи и преобразования информации; раскрывается роль и место информационных систем в народном хозяйстве, роль отечественных ученых и инженеров в развитии информационных систем и теории информации, излагаются основные определения и понятия, касающиеся основ теории прохождения сигналов в целом и теории информации в частности. Во второй и третьей главах рассмотрен сигнал как носитель информации. Приводятся характеристики основных типов детерминированных и случайных сигналов, способы их математического описания и преобразования. В четвертой и пятой главах изложены основные положения теории информации, даны понятия энтропии как

меры неопределенности и количества информации, приведены способы количественной оценки энтропии и количества информации, способы оценки скорости передачи информации и пропускной способности информационного канала, способы согласования сигналов с каналом. В шестой и седьмой главах рассматриваются критерии оценки и способы повышения эффективности и помехоустойчивости информационных систем, оценивается помехоустойчивость отдельных видов модуляции и приводятся основные положения теории эффективного и помехоустойчивого кодирования. Восьмая глава содержит основные положения теории помехоустойчивого приема. В ней рассматриваются и оцениваются различные способы обнаружения, различения и восстановления сигналов на фоне помех. Девятая глава посвящена информационной оценке автоматизированных систем контроля и управления, практическому использованию аппарата теории информации для решения отдельных задач контроля и управления. Случайный характер процессов, связанных с передачей и преобразованием сигналов, определил, что основным математическим аппаратом, используемым в книге, является аппарат теории информации и вероятностей, а также теории случайных процессов.

Авторы выражают благодарность докт. техн. наук, проф. Кошевому А. А., докт. техн. наук, проф. Расщепляеву Ю. С., докт. техн. наук, проф. Каткову Ф. А., докт. техн. наук, проф. Волкову А. А., канд. техн. наук, доц. Соколову В. А., канд. техн. наук Карпову И. М., канд. техн. наук, доц. Жуку Л. А. за критические замечания при рецензировании книги, а также докт. техн. наук, проф. Ицхоки Я. С. за идею написания учебника.

Замечания и пожелания просим направлять в Головное издательство издательского объединения «Вища школа» по адресу: 252054, Киев, 54, Гоголевская, 7.

ВВЕДЕНИЕ

Широкое внедрение автоматизированного управления немыслимо без использования средств связи, телемеханики, вычислительной техники и банков данных.

В процессе автоматизированного управления техникой, производственными процессами и отраслями народного хозяйства, а также в процессах творческой деятельности человека и общественных явлений происходит интенсивный обмен информацией между отдельными звеньями систем управления, между человеком и техникой, человеком и природой, между отдельными людьми.

Комплексная автоматизация и совершенствование электронных цифровых вычислительных машин сопровождаются резким возрастанием объема и скорости передачи и обработки информации. Одновременно повышаются требования к достоверности передачи и обработки информации.

Из всего многообразия современных технических систем можно выделить особую группу так называемых информационных систем, предназначенных для передачи, преобразования и хранения информации. К этой группе систем можно отнести: связные, телемеханические, локационные, навигационные и телевизионные системы, электронно-вычислительную и информационно-измерительную технику, автоматизированные системы управления и контроля и т. п.

Информационные системы существенно отличаются от энергетических устройств и систем (двигателей, электрических генераторов, линий электропередач, транспортных устройств, доменных печей и т. п.), основой функционирования которых являются процессы передачи и преобразования энергии. Основой функционирования информационных систем являются процессы передачи, преобразования и накопления информации. Поэтому главным критерием качества работы информационных устройств служит их способность передавать, накапливать или преобразовывать максимальное количество информации в единицу времени при допустимых искажениях и затратах, а не коэффициент полезного действия, как у энергетических установок. В информационных системах энергетические соотношения играют второстепенную роль, а сама энергия является характеристикой сигнала, который используется лишь в качестве транспортного средства.

Однако происходящее резкое возрастание потока информации в ближайшем будущем потребует значительных затрат энергии на

передачу информации. Это обусловит необходимость помимо количественной оценки информации осуществлять энергетическую оценку информационных процессов. Уже сейчас одним из критериев оценки систем космической связи является информационно-энергетический критерий, характеризующий затраты энергии сигнала на передачу единицы количества информации при данной мощности шумов, приходящейся на 1 Гц полосы пропускания канала связи [6].

По мере развития различных видов информационной техники происходила дифференциация соответствующих научных дисциплин. Возникли и развиваются такие дисциплины, как теория автоматического управления, теория автоматического контроля, радиосвязь, проводная связь, телемеханика, радиолокация, радионавигация, электронно-вычислительные машины, программирование и т. д., занимающиеся вопросами теории и синтеза отдельных видов информационных систем. Но по мере развития теории и практики этих систем определились общие основы в теории их построения, которые вылились в общетеоретическую дисциплину — кибернетику.

Академик А. Н. Колмогоров определяет кибернетику как науку «о способах восприятия, хранения, переработки и использования информации в машинах, живых организмах и их объединениях». Кибернетика устанавливает общие принципы и законы, согласно которым как живые организмы, так и некоторые машины выполняют целеустремленные действия на основе процессов передачи, преобразования и использования информации. Кибернетика прежде всего интересуется логической структурой, формализацией информационных процессов, а не их физической, биологической или какой-либо другой природой. Такое абстрагирование придает кибернетике общность, позволяет ей применять для исследования информационных процессов, независимо от того, к какой категории явление относится (биологической, физиологической, технической, экономической и т. д.), методы точных наук, весь современный математический аппарат [16].

Если основной задачей общей кибернетики является анализ информационных процессов различной природы объекта и их алгоритмизация, то основной задачей технической кибернетики является синтез информационных устройств, способных реализовать эти алгоритмы.

Предметом технической кибернетики является разработка принципов построения и теории информационных систем, реализующих алгоритмы, описывающие информационные процессы.

Теоретической основой технической кибернетики являются:

теория информации, изучающая процессы получения, преобразования, накопления и передачи информации в информационных системах;

теория логических и вычислительных машин, изучающая методы преобразования информации в соответствии с заданными алгоритмами;

теория автоматического регулирования, изучающая методы использования информации для выполнения целеустремленных действий.

В книге основное внимание уделяется теории получения, накопления, передачи и преобразования информации, анализу и синтезу информационных процессов и информационных систем.

Предполагается, что теория логических и вычислительных машин, а также теория автоматического регулирования изучаются в самостоятельных курсах.

Возникновение и развитие теории информации связано с быстрым развитием техники связи в первой половине двадцатого столетия. В этот период стали появляться работы, в которых предпринималась попытка научного обоснования основных характеристик систем связи и количественной оценки качества передачи сообщений. Еще в 1928 г. американский ученый Р. Л. Хартли предложил логарифмическую меру оценки количества информации [3]. В 1933 г. была опубликована работа советского ученого В. А. Котельникова [4], в которой впервые была сформулирована теорема о дискретном представлении сигналов с ограниченным спектром, а также дан ряд практических рекомендаций по оценке пропускной способности каналов связи. В этой работе было фактически заложено начало общей теории передачи сообщений. Исследованию вопросов увеличения пропускной способности канала связи, а также влияния помех на пропускную способность посвящена работа советского ученого Д. В. Агеева, опубликованная в 1935 году. В 1946 году опубликована работа В. А. Котельникова [5], в которой заложены основные положения теории потенциальной помехоустойчивости.

Наиболее бурное развитие теория информации получила после опубликования в 1947—1948 гг. классических работ американского математика и инженера К. Шеннона, в которых сформулированы основные положения, касающиеся количественной оценки информации. Некоторые из этих положений затем были строго доказаны советскими учеными А. Я. Хинчиным и А. Н. Колмогоровым, а также американским ученым Л. Файнстейном. Большой вклад внесли в развитие теории оптимального приема информации американские ученые Н. Винер, Д. О. Норс, Д. Мидлтон; советские ученые Л. А. Вайнштейн, В. Д. Зубаков, Л. С. Гуткин и другие. Широко известны в теории помехоустойчивого кодирования работы советских ученых К. А. Мешковского, Н. Е. Кириллова, С. И. Самойленко и Г. А. Шастовой.

Теория информации, математическим аппаратом которой является теория вероятностей и математическая статистика, превратилась к настоящему времени в строгую и достаточно универсальную науку. Эта теория широко используется для анализа процессов измерения, контроля и управления [56]. Положения этой теории все шире и шире используются при исследовании процессов мышления, познания, психологии, «согласования» сообщений с «приборами» человеческого мозга [62], а также общественных явлений.

Глава I
ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ

1.1. Основные понятия и определения

Как уже отмечалось, производственные процессы, а также процессы живой природы связаны с получением, передачей, преобразованием, накоплением, хранением и отображением информации.

В настоящее время существуют различные определения информации.

Обычно под информацией в широком смысле понимаются новые сведения об окружающем мире, которые мы получаем в результате взаимодействия с ним, приспособления к нему и изменения его в процессе приспособления.

Комитет научно-технической терминологии Академии наук СССР рекомендует следующее определение информации: информация — это сведения, являющиеся объектом хранения, передачи и преобразования [2].

Информация — это прежде всего сведения, которые должны быть использованы (сведения о состоянии природы, о состоянии и положении во времени и пространстве определенных объектов, о величине контролируемых параметров и т. п.).

В философии развернулась борьба между материалистами и идеалистами вокруг понимания природы информации. Идеалисты различных толков пытаются оторвать информацию от материи, превратить ее в некоторую духовную субстанцию или же представить в виде комплекса ощущений субъекта. В соответствии с положениями марксистской философии информация является свойством материи. Информация не является ни материей, ни энергией, но неразрывно связана с материей.

Некоторые идеалисты пытаются утверждать, что само понятие информации дает основание для примирения идеализма и материализма. Информация рассматривается ими в качестве «нейтрального элемента», не являющегося ни материей, ни сознанием, на основании чего делается вывод, что кибернетика представляет научные данные для преодоления «ограниченности» как материалистических, так и идеалистических концепций.

Понятие информации есть научное подтверждение гипотезы В. И. Ленина, что вся материя обладает свойством отражения. При этом нужно иметь в виду, что понятие «отражение» как свойство всей материи более общее, чем понятие «информация», являющееся понятием кибернетики и встречающееся только у организованных систем — машин, растений, животных и человека.

Нужно различать понятия «информация» и «сообщение». Под сообщением понимают информацию, выраженную в определенной форме и подлежащую передаче. Сообщение — это форма представления информации [2]. Примерами сообщений являются: текст телеграммы, речь оратора, показания измерительного датчика, команды управления и т. д.

Структурная схема одной из характерных информационных систем в общем случае может быть представлена в виде, изображенном на рис. 1.1. Система состоит из отправителя информации, линии связи и получателя информации.

Сообщение для передачи его в соответствующий адрес должно быть предварительно преобразовано в сигнал. Под сигналом пони-

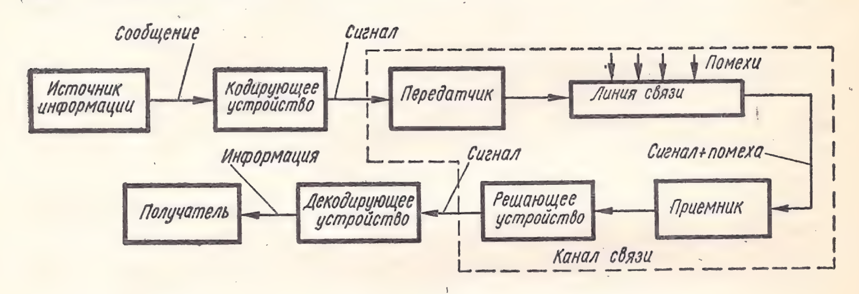


Рис. 1.1

мается изменяющаяся физическая величина, отображающая сообщение [2]. Сигнал — это материальный переносчик сообщения. Физическая среда, по которой происходит передача сигналов от передатчика к приемнику, называется линией связи.

В современной технике нашли применение электрические, электромагнитные, световые, механические, звуковые, ультразвуковые сигналы. Для передачи сообщений необходимо применять тот переносчик, который способен эффективно распространяться по используемой в системе линии связи. Например, по проводной линии связи наиболее легко проходят постоянный ток и переменные токи сравнительно невысоких частот (практически не более нескольких десятков килогерц). По радиолинии эффективно распространяются только электромагнитные колебания высоких частот (от сотен килогерц до десятков тысяч мегагерц).

Преобразование сообщений в сигналы, удобные для прохождения по линии связи, осуществляется передатчиком.

Все сообщения по характеру изменения во времени можно разделить на непрерывные и дискретные.

Непрерывные по времени сообщения отображаются непрерывной функцией времени. Дискретные по времени сообщения характеризуются тем, что поступают в определенные моменты времени и описываются дискретной функцией времени. Так как сообщения носят обычно случайный характер, то непрерывные сообщения

описываются случайной функцией времени, а дискретные сообщения — как цепь случайных событий.

Сообщения можно также разделить на непрерывные и дискретные по множеству.

Непрерывные по множеству сообщения характеризуются тем, что функция, их описывающая, может принимать непрерывное множество значений (континуум значений) в некотором интервале.

Дискретные по множеству сообщения — это сообщения, которые могут быть описаны с помощью конечного набора чисел или дискретных значений некоторой функции.

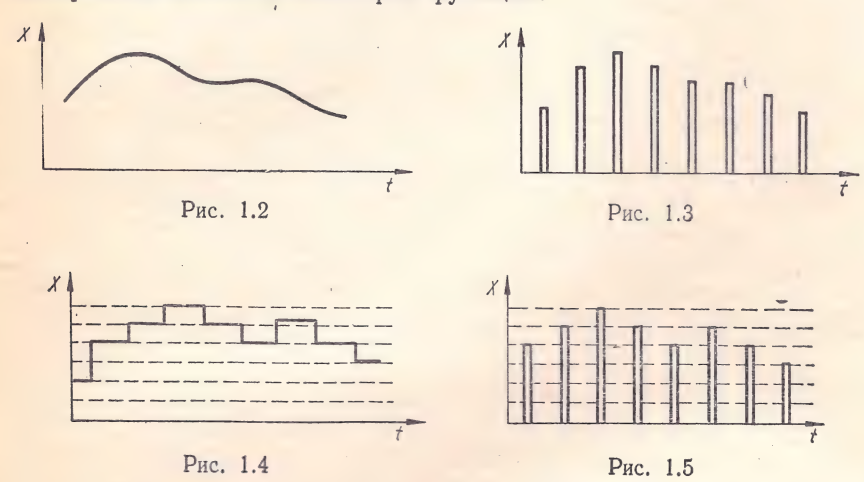


Рис. 1.2 Рис. 1.3

Рис. 1.4 Рис. 1.5

Дискретности по множеству и времени не связаны друг с другом. Поэтому возможны следующие типы сообщений:

а) непрерывные по множеству и времени, или просто непрерывные (рис. 1.2);

б) непрерывные по множеству и дискретные по времени (рис. 1.3);

в) дискретные по множеству и непрерывные по времени (рис. 1.4);

г) дискретные по множеству и времени, или просто дискретные (рис. 1.5).

В процессе преобразования дискретных сообщений в сигнал происходит кодирование сообщений. В широком смысле кодированием называют преобразование сообщений в сигнал. В узком смысле кодирование — это отображение дискретных сообщений сигналами в виде определенных сочетаний символов [2]. Устройство, осуществляющее кодирование, называется кодером (кодирующим устройством).

В передающем устройстве осуществляется воздействие на один или несколько параметров переносчика по закону, принятому при кодировании сообщений. Этот процесс называется модуляцией, а модулируемые параметры называют информативными. Иногда

применительно к радиотелеграфии в смысле модуляции употребляется термин «манипуляция».

При передаче сигналы подвергаются воздействию помех. Под помехами подразумеваются любые мешающие внешние возмущения или воздействия (атмосферные помехи, влияния посторонних источников сигналов), а также искажения сигналов в самой аппаратуре (аппаратурные помехи), вызывающие случайное отклонение принятого сообщения (сигнала) от передаваемого [2].

На приемной стороне осуществляется обратная операция декодирования, т. е. восстановление по принятому сигналу переданного сообщения. Решающее устройство, помещенное после приемника, осуществляет обработку принятого сигнала с целью наиболее полного извлечения из него информации. Декодирующее устройство (декодер) преобразует принятый сигнал к виду, удобному для восприятия получателем.

Совокупность средств, предназначенных для передачи сигнала, принято называть каналом связи [2]. Одна и та же линия связи может использоваться для передачи сигналов между многими источниками и приемниками, т. е. линия связи может обслуживать несколько каналов.

При синтезе систем передачи информации приходится решать две основные проблемы, связанные с передачей сообщений:

1) обеспечение помехоустойчивости передачи сообщений;

2) обеспечение высокой эффективности передачи сообщений.

Под помехоустойчивостью понимается способность информационной системы противостоять вредному действию помех [2]. При данных условиях, т. е. при заданной помехе, помехоустойчивость определяет верность передачи информации. Под верностью понимается мера соответствия принятого сообщения (сигнала) переданному сообщению (сигналу).

Под эффективностью системы передачи информации понимается способность системы обеспечивать передачу заданного количества информации наиболее экономичным способом. Так как передача информации сопряжена с определенными затратами мощности сигнала и времени и с использованием некоторой полосы частот, то эффективность будет характеризовать способность системы обеспечить передачу данного количества информации с наименьшими затратами мощности сигнала, времени и полосы частот.

Теория информации устанавливает критерии оценки помехоустойчивости и эффективности информационных систем, а также указывает общие пути повышения помехоустойчивости и эффективности. Повышение помехоустойчивости практически всегда сопровождается ухудшением эффективности и, наоборот, повышение эффективности отрицательно сказывается на помехоустойчивости систем.

В настоящее время теория информации успешно применяется в философии и математике, естественных и технических науках, социально-экономических науках, биологии, медицине и др. [28].

1.2. Классификация информационных систем

В информационные системы, получившие широкое распространение в народном хозяйстве, включаются средства, предназначенные для получения, преобразования, передачи, накопления, отображения и хранения информации, получаемой от человека, природы, машины, вообще от какого-либо объекта наблюдения и управления [28].

Характерной особенностью современных информационных систем является высокая степень автоматизации. Все многообразие информационных систем можно подразделить на четыре основных класса:

1. Простейшие информационные устройства, предназначенные для осуществления информационных процессов малой сложности.

К простейшим информационным устройствам относятся:
сигнализаторы и индикаторные устройства;
считывающие и регистрирующие устройства, в том числе устройства считывания магнитных и перфорационных записей;
устройства для автоматической обработки диаграмм;
читающие машины;
вычислительные устройства;
линеаризаторы, интеграторы, усреднители и устройства статистической обработки случайных величин и процессов.

2. Централизованные информационные системы (ЦИС), предназначенные для контроля и управления сложными информационными процессами.

К централизованным информационным системам относятся:
ЦИС для контроля непрерывных технологических процессов;
ЦИС для контроля дискретных технологических процессов и штучной продукции;
ЦИС для контроля производства массовой продукции.

3. Автоматические испытательные системы (АИС), предназначенные для автоматизации сложных информационных процессов.
Автоматические испытательные системы подразделяются на:
АИС для испытания сложных изделий и оборудования;
АИС для испытания радиоэлектронного оборудования;
АИС для испытания электротехнического оборудования, машин и механизмов.

4. Сложные автоматизированные информационные системы (САИС), предназначенные для осуществления централизованного контроля и управления сложными объектами, включающие автоматические и автоматизированные системы управления (АСУ).

Эти системы подразделяются на:
АСУ промышленными предприятиями (организационно-экономические и технологические);
АСУ энергосистемами;
АСУ войсками;

АСУ отраслями хозяйства, регионами и страной в целом.

Все информационные системы, кроме контроля, в большой или малой степени осуществляют управление или ведут контроль для нужд управления.

К информационным системам можно также отнести информационно-измерительные системы [63] и системы автоматического регулирования и управления, классификация которых дана в основах теории измерения, регулирования и управления.

1.3. Предмет и метод теории информации

Теория информации — это наука о получении, преобразовании, накоплении, отображении и передаче информации.

В связи с этим предметом теории информации являются процессы получения, преобразования, накопления, отображения и передачи информации в биологических, технических и смешанных системах [65].

Курс теории информации представляет собой единую научную дисциплину, основу которой составляют модели сигналов, теория случайных процессов, теория информации, теория оценки эффективности и помехоустойчивости информационных систем.

В нашей стране развитию теории информации и внедрению ее в практику придается большое значение. Без приложения теории информации практически немыслимо создание сложнейших систем управления космическими кораблями и ракетами, систем дальней и сверхдальней космической связи, систем связи и телевидения с использованием искусственных спутников Земли, сложных вычислительных и управляющих систем и т. д. [64].

Метод теории информации есть совокупность приемов исследования при анализе и синтезе сигналов и информационных систем (методы оценки информационной способности источника информации, пропускной способности систем передачи информации, информационной емкости устройств, емкости запоминающих устройств и т. д.). При этом в основе метода лежит материалистическая диалектика, объективные законы природы и общества, объективные законы действительности.

В марксистско-ленинской теории познания метод учитывает специфические закономерности деятельности мышления при отражении диалектики предмета, связывает отражение с практико-теоретическим воздействием общественного субъекта на объективный мир.

Внешний мир, окружающий человека, воздействует на него через органы чувств с усилителями или без усилителей чувствительности. Органы чувств дают человеку информацию об окружающей действительности. Получаемая информация преобразуется нервной системой и мозгом, на помощь которому может подключаться ЭВМ,

а затем, после соответствующего отбора, переработки и накопления используется человеком для обратного непосредственного воздействия на внешний мир или через орудия производства (рис. 1.6).

На этом рисунке внешний мир выступает как источник информации, а также приемник информации и воздействия. В роли усилителей чувствительности могут выступать простые и сложные измерительные системы (электронный микроскоп, радиотелескоп, космические и астрономические системы и т. д.). Возможности двигательной системы человека (исполнительных органов) могут усилить орудия труда (орудия производства).

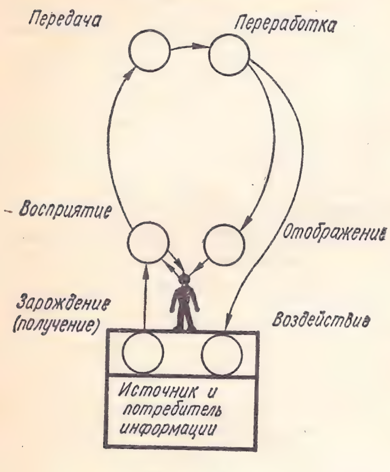


Рис. 1.6

В схеме (рис. 1.6) информация проходит следующие фазы обращения (рис. 1.7) [28].

Получение информации происходит в процессе частичного или полного «снятия» неопределенности состояния или положения объекта в пространстве и времени, в процессе познания состояния и положения объекта.

В фазе восприятия формируется образ объекта, его опознание и оценка. При этом отделяется полезная информация от шумов. В результате восприятия получается сигнал в форме, удобной для передачи и обработки.

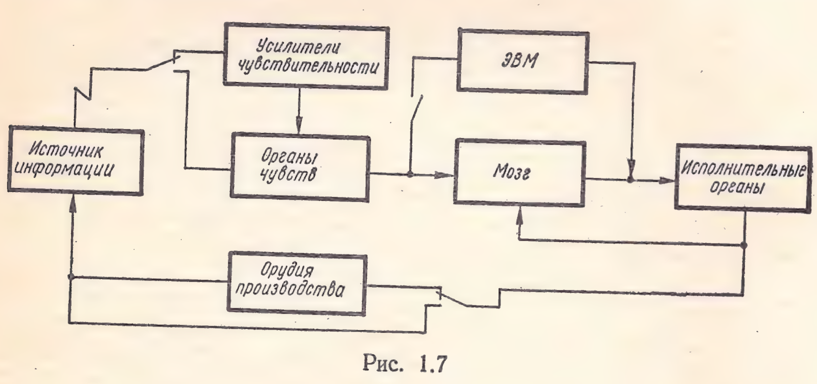


Рис. 1.7

В фазе передачи информация передается на расстояние посредством сигналов различной физической природы соответственно по механическим, гидравлическим, сейсмическим, пневматическим, акустическим, оптическим, электрическим и электромагнитным каналам. При приеме информация очищается от помех.

Переработка информации осуществляется мозгом человека или ЭВМ. Применение ЭВМ позволяет расширить возможности мозга при оценке ситуаций и принятии решений в процессах измерения, контроля и управления.

Промежуточным этапом переработки может быть запоминание информации в постоянной, долговременной или оперативной памяти.

Фаза отображения информации состоит в отображении на устройствах, способных воздействовать на органы чувств человека, качественных и количественных характеристик выходной информации.

Воздействие состоит в том, что сигналы, несущие информацию, производят регулирующие, защитные и другие действия, изменяющие состояние или положение объекта.

На практике встречаются как замкнутые, так и разомкнутые информационные системы, в которых информацией могут обмениваться человек — человек, человек — машина, машина — машина, группа машин — человек, группа операторов — машина.

С методологической точки зрения алгоритм изучения курсов «Теория информации и кодирование», «Теоретические основы информационной техники». «Теория информации» можно представить в виде структурной схемы (рис. 1.8).

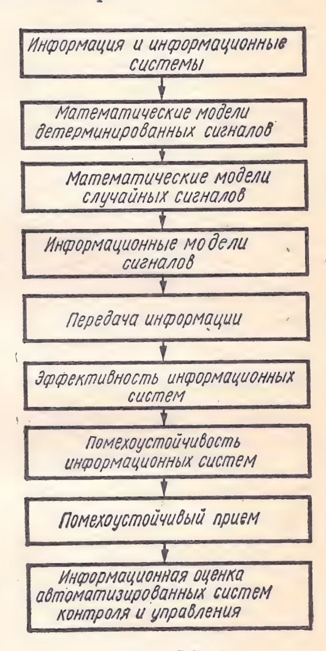


Рис. 1.8

Для успешного освоения программы курса необходимо изучить теоретические разделы, материал практических занятий и выполнить лабораторные работы.

В свою очередь, знание этого курса обеспечит изучение курсов «Теории систем», «Проектирование подсистем и звеньев АСУ», «Моделирование сложных систем», «Теория вычислительных комплексов», «Устройство и эксплуатация АСУ» и т. д.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под информационными системами?
2. Что такое информация, сообщение, сигнал?
3. Какова связь между понятиями информация, сообщение, сигнал?
4. В чем заключается процесс преобразования сообщений в сигнал?
5. Что такое линия связи, канал связи?
6. Что понимается под многоканальными системами?
7. В чем состоит предмет и метод теории информации?
8. Назвать и охарактеризовать основные фазы обращения информации, а также системы обмена информации.
9. Дать характеристику алгоритма изучения курса.

Глава II

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ

2.1. Общая характеристика сигналов

Как отмечалось, сигнал — изменяющаяся физическая величина, обеспечивающая передачу информации по линии связи. В технических информационных системах используются в большинстве случаев электрические сигналы.

Все многообразие сигналов, используемых в информационных системах, можно по своим особенностям разделить на две основные группы: детерминированные сигналы и случайные сигналы. Детерминированные сигналы характеризуются тем, что в любые моменты времени их значения являются известными величинами. Сигнал, значения которого в любые моменты времени будут случайными величинами, называется случайным.

Деление сигналов на детерминированные и случайные является условным, так как детерминированных сигналов в точном их понимании в природе нет. На практике не может быть заранее точно предсказано значение сигнала в любые моменты времени, в противном случае сигнал не нес бы полезной информации. Кроме того, любой реальный сигнал случаен в силу воздействия на него многочисленных случайных факторов.

Несмотря на это, исследование детерминированных сигналов весьма важно по двум причинам:

1) математический аппарат, используемый для анализа детерминированных сигналов, гораздо проще аппарата анализа случайных сигналов;

2) выводы, полученные в результате исследований детерминированных сигналов, могут быть во многих случаях использованы для анализа случайных сигналов.

Так как в информационных системах применяются преимущественно электрические средства передачи сообщений (проводные и радиосредства), то ограничимся рассмотрением только электрических сигналов.

Электрические сигналы являются носителями информации лишь при определенном воздействии (модуляции) на один или несколько их параметров. Такие параметры принято называть информативными или кодовыми.

Сигналы по своей структуре разделяются на непрерывные и дискретные.

Непрерывным называется сигнал, который принимает непрерывное множество значений на некотором отрезке времени и в диапазоне, ограничивающем его максимальную и минимальную величины.

Дискретным называется сигнал, принимающий конечное множество значений в определенном интервале времени и диапазоне

величин, т. е. сигнал, являющийся дискретным как по времени, так и по уровню.

Сигнал, являющийся дискретным только по времени, или только по уровню, принято называть дискретно-непрерывным. На практике из этой группы сигналов преимущественное применение нашли сигналы, квантованные только по времени. В связи с этим в дальнейшем под дискретно-непрерывным сигналом мы будем подразумевать сигналы, дискретные по времени и непрерывные по уровню.

Примерами непрерывных сигналов являются: постоянный ток и напряжение, или гармонические колебания тока (напряжения). В первом случае информативным параметром сигнала может быть только величина тока или напряжения. Во втором случае информативным параметром может быть: амплитуда, частота или фаза колебаний.

Примером дискретно-непрерывных сигналов является последовательность прямоугольных импульсов тока или напряжения. Здесь в качестве информативного параметра может быть использована: амплитуда, длительность, частота следования и фаза импульсов (временной сдвиг импульсов в каждом периоде посылки относительно опорного момента времени).

Дискретные сигналы могут представляться в виде отдельных простейших элементов (посылок) или совокупности таких элементов. Такими элементами могут быть:

однополярные импульсы напряжения (тока);

синусоидально изменяющееся напряжение (ток) определенной длительности;

периодическая последовательность импульсов напряжения (тока) определенной длительности;

пауза (отсутствие напряжения или тока).

Каждому элементу (посылке) свойственны свои качественные признаки (признаки посылок). Один или несколько из этих признаков могут быть использованы для передачи сообщений.

Для однополярных импульсов качественными признаками являются: величина, длительность и временное положение импульсов.

Для гармонической посылки качественными признаками будут: амплитуда, частота и фаза колебаний, длительность и временное положение посылки.

Для периодической последовательности импульсов качественными признаками являются: амплитуда, длительность, полярность, частота следования и фаза импульсов, длительность и временное положение посылки. Разновидностью такой посылки является периодическая последовательность импульсов высокочастотных колебаний (радиоимпульсов). Для этой посылки характерны те же качественные признаки за следующим исключением: вместо полярного признака будет фаза высокочастотных колебаний в импульсах и, кроме того, появляется дополнительный признак — частота высокочастотных колебаний.

065996

Пауза, как качественный признак, характеризуется длительностью и временным положением.

В случае использования фазы в качестве информативного параметра последняя определяется по отношению к некоторым опорным периодическим колебаниям (гармонические колебания или опорная тактовая последовательность импульсов).

В случае представления сигналов в виде совокупности элементарных посылок отдельным сообщениям соответствует определенная комбинация (кодовая комбинация) посылок. При этом должны быть приняты меры для разделения посылок в комбинациях. Применяют следующие способы разделения посылок:

1. Полярное разделение, при котором происходит чередование полярностей смежных посылок (рис. 2.1). Этот способ разделения используется для посылок в виде импульсов постоянного напряжения (тока).

2. Амплитудное разделение, при котором смежные посылки имеют различные величины амплитуд (рис. 2.2).

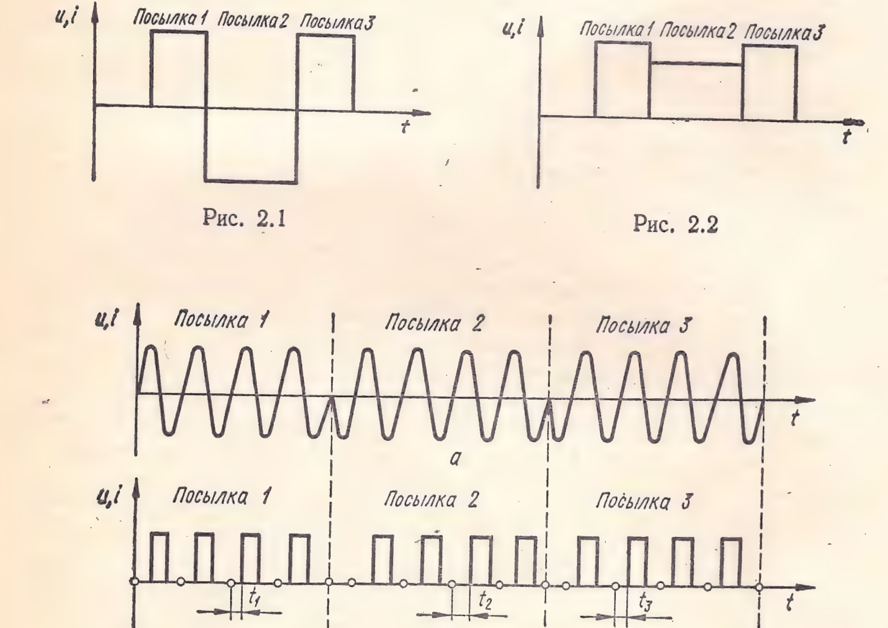


Рис. 2.1

Рис. 2.2

Рис. 2.3

3. Фазовое разделение, используемое для посылок в виде синусоидальных колебаний, при этом смежные посылки различаются фазой колебаний (обычно различие на 180°, рис. 2.3, *а*). Фазовое разделение может быть использовано и для периодических после-

довательностей однополярных импульсов. В этом случае разделение осуществляется за счет изменения временного положения импульсов в посылках (рис. 2.3, *б*).

4. Временное разделение, при котором посылки разделены друг от друга временным интервалом.

5. Частотное разделение, при котором смежные посылки различаются друг от друга частотой. Данный способ разделения применяется для посылок в виде гармонических колебаний и периодической последовательности импульсов (рис. 2.4, *а* и *б*).

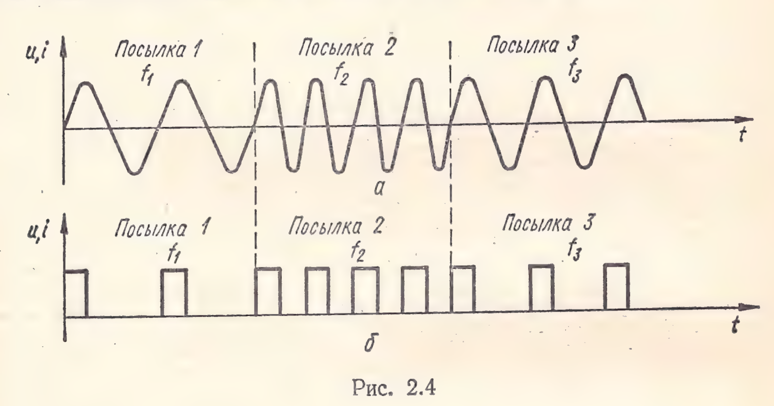


Рис. 2.4

6. Простое разделение, при котором посылки сигнала передаются одновременно по разным каналам связи.

Посылки сигнала сообщения, обладающие информативными (кодовыми) признаками, называют активными. Посылки, не обладающие такими признаками и служащие для разделения посылок, называют разделяющими. На рис. 2.5, *а* приведен пример, когда информативным признаком посылок является длительность, а разделение посылок амплитудное. Для активных посылок 1 и 3 в качестве информативного признака используется длительность. Разделяющая посылка (посылка 2) отличается от смежных информативных амплитудой.

Иногда для повышения помехоустойчивости телемеханических систем посылкам сигнала придаются дополнительные — защитные признаки. Защитные признаки в отличие от информативных могут сохранять неизменное значение для различных сигналов сообщения. На рис. 2,5, *б* приведен пример сигнала, состоящего из пяти посылок — трех активных (посылки 1, 3 и 5) и двух разделяющих (посылки 2 и 4).

Для анализа и синтеза информационных систем необходимо знать не только характеристики этих систем в виде операторов, но также и математические модели сигналов.

В зависимости от методов анализа информационных систем соответственно форме представления операторов применяются те или

иные способы представления сигналов. К основным из них следует отнести [7], [8]:

представление сигнала в виде некоторой функции времени $x(t)$;

представление сигнала в операторной форме $x(P)$;

представление сигнала в виде некоторой функции частоты.

При анализе процесса прохождения произвольного детерминированного сигнала через линейную систему или систему, близкую к линейной, удобно представлять его в виде надлежащим образом выбранной совокупности элементарных сигналов. Зная реакцию системы на элементарный сигнал, можно, пользуясь методом суперпозиции, определить реакцию системы на сигнал произвольной формы.

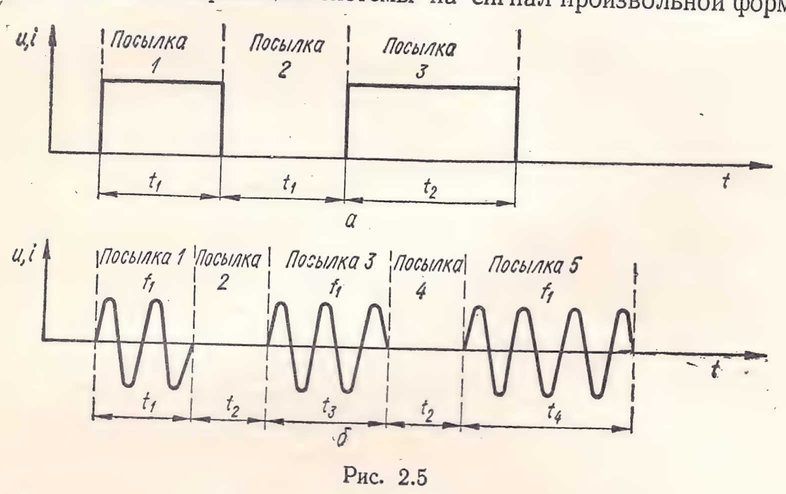


Рис. 2.5

К элементарным детерминированным сигналам относят идеальный единичный импульс, единичную функцию и синусоидальное воздействие [9].

Свойства единичного импульса (дельта-функции) определяются соотношениями

$$\delta(t-\tau)=\begin{cases}0 & \text{при } t\neq\tau;\\ \infty & \text{при } t=\tau;\end{cases} \qquad (2.1)$$

$$\int_0^t \delta(t-\tau)\,dt = 1(t-\tau), \quad 0<\tau<t,$$

где $\delta(t-\tau)$ — дельта-функция; t — время; τ — момент действия импульса; $1(t-\tau)$ — единичная функция.

Следовательно, единичный импульс — это идеализированный сигнал, характеризуемый бесконечно малой длительностью, бесконечно большим уровнем и площадью, равной единице.

Свойства единичной функции определяются соотношениями:

$$1(t-\tau)=\begin{cases}0 & \text{при } t<\tau;\\ 1 & \text{при } t\geqslant\tau.\end{cases} \qquad (2.2)$$

Таким образом, единичная функция — это временная функция, которая при любом $t < \tau$ тождественно равна нулю, а при любом $t \geqslant \tau$ равна 1. Единичная функция выражает скачкообразное изменение в момент времени $t = \tau$ величины от значения 0 до значения 1.

При анализе многих типов технических информационных систем, в особенности при исследовании пропускной способности и селективности систем, широко используются синусоидальные элементарные сигналы. В связи с этим большой интерес имеет частотное представление сигнала, т. е. представление модели сигнала в виде совокупности элементарных синусоидальных сигналов.

2.2. Частотное представление детерминированных сигналов

В частотном виде могут представляться как периодические, так и непериодические детерминированные сигналы.

Необходимо заметить, что в реальных условиях периодические сигналы не существуют, так как идеальный периодический сигнал бесконечен во времени, в то время как всякий реальный сигнал имеет начало и конец. Однако во многих случаях конечностью времени действия сигнала можно пренебречь и для его анализа допустимо использовать аппарат, пригодный для идеальных периодических сигналов.

1. Периодические сигналы

Рассмотрим сигнал, выражаемый произвольной периодической функцией времени $x(t)$ (рис. 2.6).

Известно, что всякая периодическая функция, удовлетворяющая условиям Дирихле[1], может быть представлена в виде бесконечной в общем случае суммы гармонических составляющих — рядом Фурье.

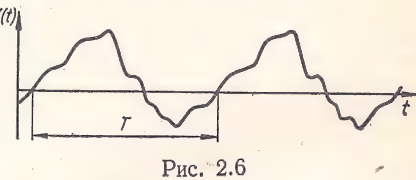


Рис. 2.6

Известны две формы разложения в ряд Фурье: тригонометрическая и комплексная. Тригонометрическая форма разложения выражается в виде

$$x(t) = \frac{1}{2} A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega_0 t - \varphi_k), \qquad (2.3)$$

где $\frac{1}{2} A_0$ — постоянная составляющая функции $x(t)$; $A_k \cos(k\omega_0 t - \varphi_k)$ — k-я гармоническая составляющая; A_k, $k\omega_0$, φ_k — амплитуда, частота и начальная фаза k-й гармонической составляющей; $\omega_0 =$

[1] Условие Дирихле заключается в следующем: функция $x(t)$ должна быть ограниченной, кусочно-непрерывной и иметь на протяжении периода конечное число экстремальных значений.

$=\frac{2\pi}{T}$ — частота основной (первой) гармоники; T — период колебаний.

В математическом отношении удобнее оперировать комплексной формой ряда Фурье, представляемой в виде

$$x(t)=\frac{1}{2}\sum_{k=-\infty}^{\infty}\dot{A}_k\exp\{jk\omega_0 t\}, \qquad (2.4)$$

где $\dot{A}_k=A_k\exp\{-j\varphi_k\}$ — комплексная амплитуда гармонической составляющей частоты $\omega_k=k\omega_0$.

Суммы, определяемые выражениями (2.3) и (2.4), будут тождественными при выполнении условий

$$A_{-k}=A_k;\quad \varphi_{-k}=\varphi_k;\ \varphi=0. \qquad (2.5)$$

При этом модуль комплексной амплитуды будет равен амплитуде соответствующей гармонической составляющей, а аргумент равен начальной фазе составляющей.

Комплексная амплитуда определяется с помощью формулы

$$\dot{A}_k=\frac{2}{T}\int_{t_1}^{t_1+T}x(t)\exp\{-jk\omega_0 t\}\,dt. \qquad (2.6)$$

Совокупность амплитуд и соответствующих частот гармоник принято называть спектром амплитуд. Совокупность начальных фаз и соответствующих частот гармоник называют спектром фаз.

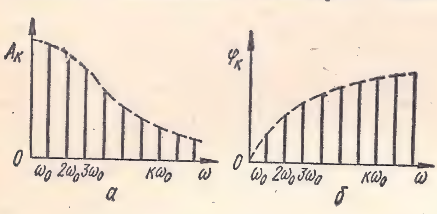


Рис. 2.7

Спектр амплитуд и спектр фаз однозначно определяют сигнал. Однако для многих практических задач достаточно ограничиться рассмотрением только спектра амплитуд.

На рис. 2.7, *а* и *б* даны графические изображения спектра амплитуд и спектра фаз периодического сигнала. Отдельные спектральные составляющие в графическом изображении спектра амплитуд называются спектральными линиями.

Характерной особенностью спектра периодического сигнала является его прерывистость (дискретность). Расстояние между соседними спектральными линиями одинаковое и равно частоте основной гармоники.

Рассмотрим спектры периодических сигналов при различных видах модуляции.

а) Амплитудно-модулированный гармонический сигнал.

$$x(t)=A(t)\sin(\omega_0 t+\varphi_0).$$

При амплитудной модуляции амплитуда изменяется по определенному закону

$$A(t) = A_0 + \delta A f(t),$$

где A_0 — постоянная составляющая амплитуд; δA — наибольшее изменение амплитуды при модуляции; $f(t)$ — нормированная функция (изменяющаяся в пределах от -1 до $+1$).

Так как модулируемый параметр сигнала (в данном случае амплитуда) является непосредственным переносчиком информации, то функция $f(t)$ выражает закон изменения во времени передаваемого сообщения.

Амплитудно-модулированный гармонический сигнал как функция времени в общем случае имеет вид

$$x(t) = A_0 [1 + m_A f(t)] \sin(\omega_0 t + \varphi_0), \quad (2.7)$$

где $m_A = \frac{\delta A}{A_0}$ — глубина амплитудной модуляции.

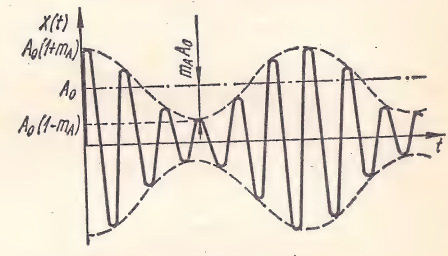


Рис. 2.8

Рассмотрим частный случай, когда функция $f(t)$ изменяется по гармоническому закону (рис. 2.8)

$$f(t) = \cos \Omega t, \text{ причем } \Omega \ll \omega_0.$$

Тогда выражение (2.7) примет вид

$$x(t) = A_0 [1 + m_A \cos \Omega t] \sin(\omega_0 t + \varphi_0) =$$

$$= A_0 \sin(\omega_0 t + \varphi_0) + \frac{A_0 m_A}{2} \sin[(\omega_0 - \Omega) t + \varphi_0] +$$

$$+ \frac{A_0 m_A}{2} \sin[(\omega_0 + \Omega) t + \varphi_0]. \quad (2.8)$$

Как видно из выражения (2.8), спектр сигнала, изображенного на рис. 2.8, состоит из трех гармонических составляющих: несущей с частотой ω_0 и двух боковых — нижней с частотой $\omega_0 - \Omega$ и верхней с частотой $\omega_0 + \Omega$ (рис. 2.9). Ширина спектра сигнала $\Delta\omega = 2\Omega$.

Если модуляция гармонического сигнала производится по более сложному закону, причем спектр огибающей амплитуды находится в диапазоне частот от $\Omega_{мин}$ до $\Omega_{макс}$, то можно показать, что в спектре амплитудно-модулированного сигнала вместо двух боковых частот будут две боковые полосы частот (рис. 2.10). Причем, нижняя боковая полоса частот будет находиться в пределах $\omega_0 - \Omega_{макс}$, $\omega_0 - \Omega_{мин}$, а верхняя боковая полоса частот — в пределах $\omega_0 + \Omega_{макс}$, $\omega_0 + \Omega_{мин}$.

Для неискаженной передачи такого сигнала канал связи должен обладать полосой пропускания частот, равной $2\Omega_{макс}$, т. е. полоса пропускания канала связи должна быть вдвое больше наивысшей частоты спектра модулирующего сигнала.

Для уменьшения полосы частот модулированного сигнала, а следовательно, и требуемой полосы пропускания канала связи иногда используется так называемая однополосная передача. При такой передаче с помощью фильтров производится подавление несущей частоты и частот одной из боковых полос, т. е. передача производится в полосе частот $\Omega_{макс} - \Omega_{мин}$. Это позволяет более чем в 2 раза сократить требуемую полосу канала связи.

б) Частотно-модулированный гармонический сигнал.

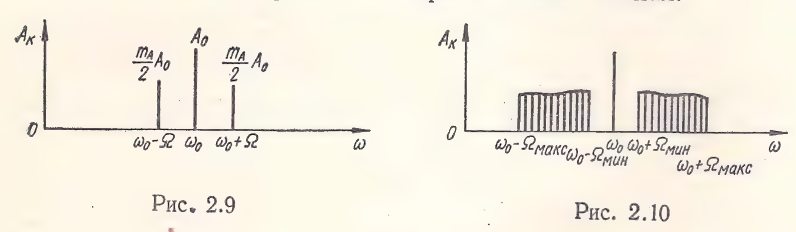


Рис. 2.9

Рис. 2.10

При частотной модуляции частота сигнала изменяется в общем виде по закону

$$\omega(t) = \omega_0 [1 + m_\omega f(t)], \tag{2.9}$$

где ω_0 — постоянная составляющая частоты; $m_\omega = \frac{\delta\omega}{\omega_0}$ — глубина частотной модуляции; $\delta\omega$ — наибольшее изменение частоты при модуляции (девиация частот); $f(t)$ — закон модуляции.

Частотно-модулированный сигнал можно представить в виде

$$x(t) = A_0 \sin\varphi(t) = A_0 \sin\left[\int_0^t \omega(t)\,dt + \varphi_0\right] =$$

$$= A_0 \sin\left[\omega_0 t + \omega_0 m_\omega \int_0^t f(t)\,dt + \varphi_0\right], \tag{2.10}$$

где $\varphi(t)$ — текущая фаза гармонического сигнала; φ_0 — начальная фаза сигнала.

В частном случае, когда модуляция частоты осуществляется по гармоническому закону, частотно-модулированный сигнал может быть записан в виде

$$x(t) = A_0 \sin\left[\omega_0 t + \omega_0 m_\omega \int_0^t \cos\Omega t\,dt + \varphi_0\right] = A_0 \sin(\omega_0 t + \beta \sin\Omega t) =$$

$$= A_0 \sin\omega_0 t \cos[\beta\sin\Omega t] + A_0 \cos\omega_0 t \sin[\beta \sin\Omega t], \tag{2.11}$$

где $\beta = \frac{\delta\omega}{\Omega}$ — индекс частотной модуляции.

Сигнал такого вида представляется следующим образом:

$$x(t) = A_0 \{ \dot{J}_0(\beta) \sin \omega_0 t + \sum_{k=1}^{\infty} \dot{J}_k(\beta) [\sin(\omega_0 + k\Omega) t + $$
$$+ (-1)^k \sin(\omega_0 - k\Omega) t] \}, \qquad (2.12)$$

где $\dot{J}_0(\beta)$— функция Бесселя нулевого порядка; $\dot{J}_k(\beta)$ — функция Бесселя k-го порядка.

Спектр амплитуд модулированного по частоте сигнала приведен на рис. 2.11. Спектр дискретный, состоит из колебаний с несущей частотой ω_0 и бесконечного количества верхних и нижних боковых составляющих с частотами $\omega \pm k\Omega$.

Рис. 2.11

Распределение амплитуд гармонических составляющих зависит от аргумента β функций Бесселя, т. е. от индекса частотной модуляции.

При очень малом индексе частотной модуляции ($\beta \to 0$) спектр частотно-модулированного сигнала практически не отличается от спектра амплитудно-модулированного сигнала, т. е. он состоит всего из трех гармонических составляющих: колебаний с несущей частотой ω_0 и двух боковых с частотами $\omega_0 - \Omega$ и $\omega_0 + \Omega$. В этом случае полоса пропускания канала связи должна быть равна 2Ω.

С ростом индекса модуляции β увеличивается вес боковых составляющих и соответственно требуемая полоса пропускания канала связи. Если ограничиться в спектре только составляющими, амплитуда которых составляет не менее 5—10% от амплитуды несущей A_0 до модуляции, то ширина спектра этих составляющих, а следовательно, и требуемая полоса канала связи будут составлять $2(\delta\omega + \Omega)$.

При больших индексах модуляции ($\beta \gg 1$) ширина спектра сигнала практически равна удвоенной величине девиации частоты ($2\delta\omega$) и не зависит от частоты модулирующего сигнала Ω.

в) Фазомодулированный гармонический сигнал

$$x(t) = A_0 \sin[\omega_0 t + \varphi(t)].$$

При фазовой модуляции фаза гармонического сигнала изменяется по закону модулирующего сигнала $f(t)$

$$\varphi(t) = \varphi_0 [1 + m_\varphi f(t)], \qquad (2.13)$$

где $m_\varphi = \frac{\delta\varphi}{\varphi_0}$ — глубина фазовой модуляции; $\delta\varphi$ — наибольшее изменение фазы при модуляции (индекс фазовой модуляции).

Сигнал с фазовой модуляцией можно представить в следующем виде:

$$x(t) = A_0 \sin[\omega_0 t + \varphi_0 m_\varphi f(t) + \varphi_0]. \qquad (2.14)$$

Так как мгновенное значение частоты является производной от фазы по времени

$$\omega = \frac{d\varphi}{dt},$$

то при фазовой модуляции оно может быть представлено в виде

$$\omega(t) = \omega_0 + \varphi_0 m_\varphi \frac{df(t)}{dt}.$$

Таким образом, фазомодулированный сигнал эквивалентен сигналу с частотной модуляцией с модулирующей функцией $\frac{df(t)}{dt}$.

Для частного случая, когда модулирующий сигнал является гармоническим, полная фаза фазомодулированного сигнала определяется равенством

$$\varphi(t) = \omega_0 t + \varphi_0 m_\varphi \cos \Omega t + \varphi_0, \quad (2.15)$$

а фазомодулированное колебание опишется выражением

$$x(t) = A_0 \sin[\omega_0 t + \varphi_0 m_\varphi \cos \Omega t + \varphi_0]. \quad (2.16)$$

Мгновенное значение частоты фазомодулированного колебания

$$\omega(t) = \frac{d\varphi(t)}{dt} = \omega_0 - \delta\varphi\Omega \sin \Omega t = \omega_0 - \omega_в \sin \Omega t,$$

где $\omega_в = \delta\varphi\Omega$ — девиация частоты фазомодулированного колебания.

Сравнение выражений (2.16) и (2.11) показывает, что при гармоническом модулирующем сигнале выражение, описывающее частотно-модулированное колебание, отличается от такого же для фазомодулированного колебания только фазой гармонической функции, определяющей изменение полной фазы высокочастотного колебания. В связи с этим по внешнему виду сигнала нельзя заключить, как модулирован сигнал — по частоте или фазе.

Однако при изменении частоты модуляции Ω проявляется различие между частотной и фазовой модуляциями.

При частотной модуляции величина девиации частоты $\omega_в$ зависит только от амплитуды модулирующего сигнала и не зависит от частоты модуляции Ω. Величина же индекса β убывает с ростом частоты Ω.

При фазовой модуляции величина индекса модуляции определяется выражением $\beta = \frac{\omega_в}{\Omega} = \delta\varphi$, т. е. зависит только от амплитуды модулирующего сигнала и не зависит от частоты модуляции.

В общем случае спектр колебания, модулированного по фазе одной частотой Ω, выражается аналитически так же, как и спектр частотно-модулированного сигнала с такой же частотой модуляции Ω.

Спектр фазомодулированного сигнала состоит из колебаний с несущей частотой ω_0 и бесконечного количества верхних и нижних боковых составляющих с частотами $\omega_0 \pm k\Omega$. Амплитуды

боковых составляющих определяются только индексом модуляции β. При очень малом значении индекса фазовой модуляции β спектр фазомодулированного сигнала состоит из несущей частоты ω_0 и двух боковых частот $\omega_0 + \Omega$ и $\omega_0 - \Omega$. Ширина спектра равна 2Ω. При $\beta \gg 1$ практическая ширина спектра сигнала равна удвоенной девиации частоты $\Delta\omega = 2\omega_в$, но зависит как от амплитуды, так и от частоты модулирующего сигнала, в то время как при частотной модуляции ширина спектра в данном случае зависит только от амплитуды модулирующего сигнала.

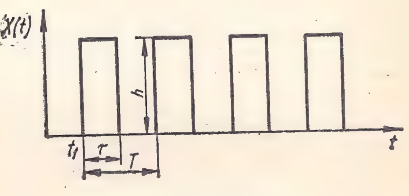


Рис. 2.12

Практическая ширина спектра частотно-модулированного и фазомодулированного сигнала примерно в β раз шире спектра амплитудно-модулированного сигнала при такой же частоте модуляции.

г) Периодическая последовательность прямоугольных импульсов длительностью τ, амплитудой h, с периодом следования T (рис. 2.12).

Функция $x(t)$, описывающая такой сигнал, может быть представлена следующим образом:

$$x(t) = \begin{cases} h \text{ при } t_1 \leqslant t \leqslant t_1 + \tau; \\ 0 \text{ при } t_1 + \tau < t < t_1 + T. \end{cases} \qquad (2.17)$$

Функция $x(t)$ может быть представлена рядом Фурье

$$x(t) = \frac{1}{2} \sum_{k=-\infty}^{\infty} \dot{A}_k \exp\{jk\omega_0 t\} = \frac{1}{2} A_0 + \sum_{k=1}^{\infty} (\dot{A}_k \exp\{jk\omega_0 t\} + \dot{A}_k \exp\{-jk\omega_0 t\}, \qquad (2.18)$$

где $\omega_0 = \frac{2\pi}{T}$; $\dot{A}_k$ — комплексные амплитуды k-й гармоники; $\frac{1}{2} A_0$ — постоянная составляющая сигнала.

$$\dot{A}_k = A_k \exp\{j\varphi_k\}; \quad \dot{A}_{-k} = A_k \exp\{-j\varphi_k\}.$$

Согласно (2.6) комплексная амплитуда

$$\dot{A}_k = \frac{2}{T} \int_{t_1}^{t_1+T} x(t) \exp\{-jk\omega_0 t\}\, dt = \frac{2}{T} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} h \exp\{-jk\omega_0 t\}\, dt =$$

$$= \frac{2h}{T} \left. \frac{\exp\{-jk\omega_0 t\}}{-jk\omega_0} \right|_{-\frac{\tau}{2}}^{\frac{\tau}{2}} = 2\frac{\tau h}{T} \frac{\sin\frac{k\omega_0\tau}{2}}{\frac{k\omega_0\tau}{2}}. \qquad (2.19)$$

Если длительность импульсов $\tau = \frac{T}{2}$, т. е. частота основной гармоники $\omega_0 = \frac{2\pi}{T} = \frac{\pi}{\tau}$, то комплексная амплитуда

$$\dot{A}_k = \begin{cases} \frac{2h}{k\pi} & \text{при } k \text{ — нечетном;} \\ 0 & \text{при } k \text{ — четном.} \end{cases} \tag{2.22}$$

Таким образом, в составе спектра сигнала будут только нечетные составляющие.

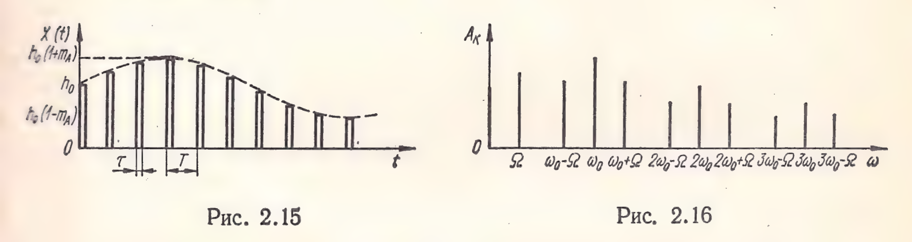


Рис. 2.15

Рис. 2.16

Огибающая спектра амплитуд описывается уравнением

$$A(\omega) = \frac{2h}{\omega\tau}.$$

Спектр амплитуд показан на рис. 2.14.

д) Амплитудно-модулированная периодическая последовательность импульсов (рис. 2.15) с длительностью импульсов τ и периодом следования T.

Амплитуда импульсов модулируется по гармоническому закону

$$h(t) = h_0[1 + m_A \sin(\Omega t + \varphi_0)]. \tag{2.23}$$

Подставив (2.23) в выражение (2.20) для немодулированной периодической последовательности импульсов, получим выражение для амплитудно-модулированной последовательности прямоугольных импульсов

$$x(t) = \frac{\tau h_0}{T}[1 + m_A \sin(\Omega t + \varphi_0)] \left[1 + 2 \sum_{k=1}^{\infty} \frac{\sin \frac{k\omega_0\tau}{2}}{\frac{k\omega_0\tau}{2}} \cos k\omega_0 t \right]. \tag{2.24}$$

После определенных тригонометрических преобразований получим

$$x(t) = \frac{\tau h_0}{T} \left[1 + 2 \sum_{k=1}^{\infty} \frac{\sin \frac{k\omega_0\tau}{2}}{\frac{k\omega_0\tau}{2}} \cos k\omega_0 t \right] + m_A \frac{\tau h_0}{T} \sin(\Omega t + \varphi_0) +$$

$$+ m_A \frac{\tau h_0}{T} \sum_{k=1}^{\infty} \frac{\sin \frac{k\omega_0\tau}{2}}{\frac{k\omega_0\tau}{2}} \sin [(k\omega_0 + \Omega) t + \varphi_0] +$$

$$+ m_A \frac{\tau h_0}{T} \sum_{k=1}^{\infty} \frac{\sin \frac{k\omega_0\tau}{2}}{\frac{k\omega_0\tau}{2}} \sin [(k\omega_0 - \Omega) t - \varphi_0]. \qquad (2.25)$$

Сравнивая полученное выражение с разложением (2.20) для периодической последовательности немодулированных импульсов, видим, что в результате амплитудной модуляции импульсов в спектре, кроме постоянной составляющей и составляющих с частотами, кратными основной частоте ω_0, появились составляющая с частотой модуляции Ω, а также по две боковые составляющие—нижняя $k\omega_0 - \Omega$ и верхняя $k\omega_0 + \Omega$ — около каждой k-й гармоники основной частоты.

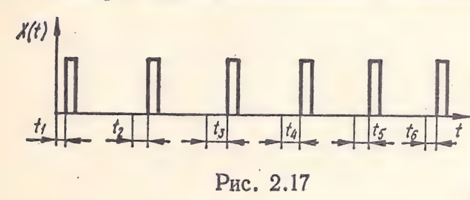


Рис. 2.17

Спектр амплитудно-модулированной последовательности импульсов показан на рис. 2.16.

Появление боковых частот $k\omega_0 \pm \Omega$ практически не сказывается на ширине спектра сигнала, так как частота модуляции $\Omega \ll \omega_0$. В связи с этим практическая ширина спектра амплитудно-модулированной последовательности импульсов такая же, как и соответствующей последовательности немодулированных импульсов, т. е. определяется длительностью импульсов.

е) Фазомодулированная последовательность импульсов (рис. 2.17).

В случае использования фазоимпульсной модуляции (ФИМ) передача сообщений осуществляется с помощью последовательности импульсов постоянной амплитуды.

Разложение такого сигнала в ряд Фурье для случая, когда модуляция осуществляется по гармоническому закону с частотой Ω, представляется в виде

$$x(t) = h_0 \frac{\tau}{T} + 2h_0 \frac{\theta}{T} \sin \Omega \frac{\tau}{2} \cos \Omega \left(t - \frac{\tau}{2}\right) + \sum_{k=1}^{\infty} M_k \cos k\omega_0 t +$$

$$+ \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} N_{kn} \cos \left[(k\omega_0 - n\Omega) t + k\Omega \frac{\tau}{2}\right] + \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} Q_{kn} \cos \times$$

$$\times \left[(k\omega_0 + n\Omega) t - n\Omega \frac{\tau}{2}\right], \qquad (2.26)$$

где θ — наибольшее смещение импульсов относительно тактовых точек при модуляции

$$M_k = \frac{2h_0}{k\pi} J_0(k\omega_0\theta) \sin k\omega_0 \frac{\tau}{2};$$

$$N_{kn} = \frac{2h_0}{k\pi} J_n(k\omega_0\theta) \sin(k\omega_0 + n\Omega) \frac{\tau}{2};$$

$$Q_{kn} = \frac{2h_0}{k\pi} J_n(k\omega_0\theta) \sin(k\omega_0 - n\Omega) \frac{\tau}{2},$$

где $J_0(k\omega_0\theta)$ — функция Бесселя нулевого порядка; $J_n(k\omega_0\theta)$ — функция Бесселя n-го порядка.

В отличие от сигналов с амплитудно-импульсной модуляцией в спектре сигнала с фазоимпульсной модуляцией около составляющих основной частоты и кратных ей имеются не по две боковые частоты, а по две дискретных полосы боковых гармоник с частотами $k\omega_0 \pm n\Omega$. Однако практическая ширина спектра частот сигнала при фазоимпульсной модуляции такая же, как и при амплитудно-импульсной модуляции.

Не приводя довольно сложных выражений разложения в ряд Фурье последовательности прямоугольных импульсов с широтной и частотной модуляцией, заметим, что при этих видах модуляции практическая ширина спектра определяется в основном длительностью импульсов.

2. Непериодические сигналы

Всякий непериодический сигнал (рис. 2.18) можно рассматривать как периодический, период изменения которого равен бесконечности. В связи с этим рассмотренный ранее спектральный анализ периодических процессов может быть обобщен и на непериодический сигнал.

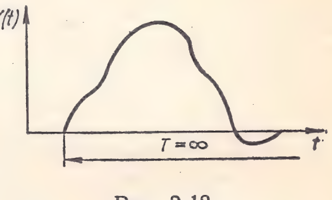


Рис. 2.18

Рассмотрим, как будет изменяться спектр периодического сигнала при неограниченном увеличении периода изменения сигнала.

При увеличении периода T интервалы между смежными частотами в спектре сигнала и амплитуды спектральных составляющих уменьшаются и в пределе при $T \to \infty$ становятся бесконечно малыми величинами. При этом ряд Фурье, отображающий спектральное разложение периодического сигнала, преобразуется в интеграл Фурье, отображающий спектральное разложение непериодического сигнала.

Комплексная форма интеграла Фурье имеет вид

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) \exp\{ j\omega t \} d\omega, \tag{2.27}$$

где $S(j\omega) = |S(j\omega)| \exp\{j\varphi(\omega)\}$ — спектральная плотность; $|S(j\omega)| = S(\omega)$ — амплитудно-частотная характеристика сигнала; $\varphi(\omega)$ — фазочастотная характеристика сигнала.

Выражение (2.27) называют формулой обратного преобразования Фурье.

Представление непериодической функции интегралом Фурье возможно при выполнении следующих условий:

1) функция $x(t)$ удовлетворяет условиям Дирихле;

2) функция $x(t)$ абсолютно интегрируема[1], т. е.

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty.$$

Таким образом, спектр непериодического сигнала в отличие от спектра периодического сигнала является сплошным и представляет собой сумму бесконечного числа гармонических составляющих с бесконечно малыми амплитудами.

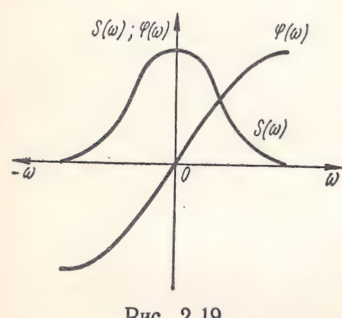


Рис. 2.19

Амплитуды гармонических составляющих, исходя из (2.27), могут быть представлены в виде

$$dA = \frac{1}{2\pi} S(j\omega)\, d\omega,$$

откуда спектральная плотность будет определяться выражением

$$S(j\omega) = 2\pi \frac{dA}{d\omega}.$$

Спектральная плотность связана с временной функцией сигнала через прямое преобразование Фурье

$$S(j\omega) = \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t)\}\, dt. \qquad (2.28)$$

Спектральная плотность однозначно отображает непериодический сигнал и удовлетворяет условиям (рис. 2.19):

1. $\lim\limits_{\omega \to \infty} S(\omega) = 0.$

2. Модуль спектральной плотности является четной, а аргумент — нечетной функцией частоты, т. е.

$$S(\omega) = S(-\omega); \qquad \varphi(\omega) = -\varphi(-\omega).$$

3. Сопоставление спектров периодических сигналов и соответствующих непериодических сигналов

Рассмотрим периодический сигнал $x_n(t)$ с периодом изменения T (рис. 2.20, *а*) и непериодический сигнал $x(t)$ (рис. 2.20, *б*), который в интервале t_1, $(t_1 + T)$ совпадает с периодическим сигна-

[1] Этим условиям удовлетворяет практически любой реальный сигнал.

лом, а вне этого интервала равен нулю. Комплексная амплитуда гармоник для разложения периодического сигнала и спектральная плотность для разложения непериодического сигнала соответственно

$$\dot{A}_k = \frac{2}{T} \int\limits_{t_1}^{t_1+T} x_n(t) \exp\{-jk\omega t\}\, dt; \qquad (2.29)$$

$$S(j\omega) = \int\limits_{-\infty}^{\infty} x(t) \exp\{-j\omega t\}\, dt = \int\limits_{t_1}^{t_1+T} x_n(t) \exp\{-j\omega t\}\, dt. \qquad (2.30)$$

Из сравнения выражений (2.29) и (2.30) можно заключить, что для каждой частоты спектра периодического сигнала справедливы равенства

$$\dot{A}_k = \frac{2}{T} S(jk\omega_0);$$

$$A_k = |\dot{A}_k| = \frac{2}{T} S(k\omega_0); \qquad (2.31)$$

$$A_0 = |\dot{A}_0| = \frac{2}{T} S(0).$$

Отсюда следует, что спектр непериодического сигнала, заданного в интервале $t_1 \leqslant t \leqslant t_1 + T$, с точностью до постоянного мно-

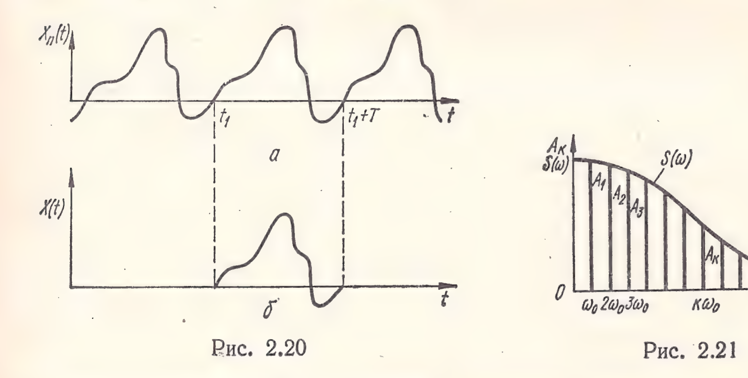


Рис. 2.20

Рис. 2.21

жителя совпадает с огибающей спектра периодического сигнала в интервале $t_1 \leqslant t \leqslant t_1 + T$ описываемого функцией такого же вида, что и непериодический сигнал (рис. 2.21).

Таким образом, спектр непериодического сигнала может быть определен по известному спектру соответствующего периодического сигнала и, наоборот, спектр периодического сигнала может быть найден по известному спектру соответствующего непериодического сигнала.

4. Энергетическое толкование спектра сигнала

Рассмотрим распределение мощности в спектре периодического сигнала. Для этого предположим, что сигнал представляет собой ток $i(t)$, протекающий по резистору R (рис. 2.22) и описываемый сложной периодической функцией времени с периодом изменения t.

Средняя мощность, выделяемая на резисторе R,

$$P_{\text{ср}} = \frac{1}{T}\int_0^T i^2(t)\,dt = rI^2,$$

где $I^2 = \frac{1}{T}\int_0^T i^2(t)\,dt$ — квадрат действующего значения тока.

Представив ток $i(t)$ рядом Фурье (2.4), получим следующее выражение для квадрата действующего значения тока:

$$I^2 = \frac{1}{T}\int_0^T \left[\frac{1}{2}A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega_0 t - \varphi_k)\right]^2 dt = \frac{1}{4T}\int_0^T A_0^2\,dt +$$

$$+ \frac{A_0}{T}\sum_{k=1}^{\infty}\int_0^T A_k \cos(k\omega_0 t - \varphi_k)\,dt + \frac{1}{T_l}\sum_{\substack{l=0\\ i\neq l}}\int_0^T A_i A_l \cos(i\omega_0 t - \varphi_i)\,\times$$

$$\times \cos(l\omega_0 t - \varphi_l)\,dt + \frac{1}{T}\sum_{k=1}^{\infty}\int_0^T A_k^2 \cos^2(k\omega_0 t - \varphi_k)\,dt = \frac{A_0^2}{4} + \frac{1}{2}\sum_{k=1}^{\infty} A_k^2.$$

Следовательно, средняя мощность

$$P_{\text{ср}} = r\left(\frac{A_0}{2}\right)^2 + \frac{r}{2}\sum_{k=1}^{\infty} A_k^2. \qquad (2.32)$$

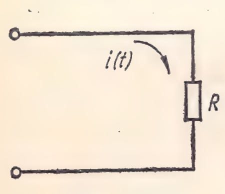


Рис. 2.22

Таким образом, средняя мощность, выделяемая сложным периодическим током в резисторе, равна сумме средних мощностей, выделяемых в этом резисторе отдельными гармониками тока и его постоянной составляющей.

Рассмотрим теперь распределение энергии в спектре непериодического сигнала.

Энергия, выделяемая сигналом в резисторе в один Ом, определяется выражением

$$W = \int_{-\infty}^{\infty} [x(t)]^2\,dt. \qquad (2.33)$$

Если функция $x(t)$ описывает непериодический ток, то формула (2.33) определяет энергию, выделяемую током в резисторе в один Ом.

Для определения распределения энергии по спектру непериодического сигнала выразим энергию W через модуль спектральной плотности сигнала $S(\omega)$ [10].

Квадрат модуля спектральной плотности можно представить в виде

$$[S(\omega)]^2 = S(j\omega)\, S(-j\omega),$$

где $S(-j\omega)$ — комплексно-сопряженная функция для спектральной плотности $S(j\omega)$.

Согласно (2.28)

$$S(-j\omega) = \int\limits_{-\infty}^{\infty} x(t) \exp\{ j\omega t \}\, dt.$$

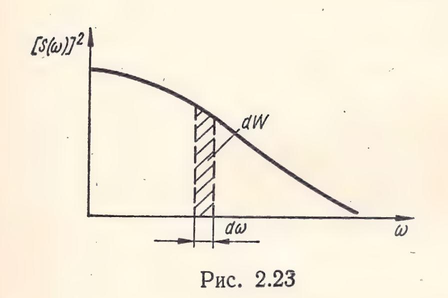


Рис. 2.23

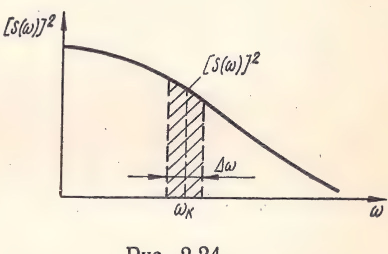


Рис. 2.24

Интеграл от квадрата модуля спектральной плотности

$$\int\limits_{-\infty}^{\infty} [S(\omega)]^2\, d\omega - \int\limits_{-\infty}^{\infty} S(j\omega)\, S(-j\omega)\, d\omega =$$

$$= \int\limits_{-\infty}^{\infty} S(j\omega) \int\limits_{-\infty}^{\infty} x(t) \exp\{ j\omega t \}\, dt d\omega. \qquad (2.34)$$

Изменив в (2.34) порядок интегрирования, получим

$$\int\limits_{-\infty}^{\infty} [S(\omega)]^2\, d\omega = \int\limits_{-\infty}^{\infty} x(t) [\int\limits_{-\infty}^{\infty} S(j\omega) \exp\{ j\omega t \}\, d\omega]\, dt = 2\pi \int\limits_{-\infty}^{\infty} [x(t)]^2\, dt.$$

Таким образом, энергия сигнала

$$W = \int\limits_{-\infty}^{\infty} [x(t)]^2\, dt = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} [S(\omega)]^2\, d\omega = \frac{1}{\pi} \int\limits_{0}^{\infty} [S(\omega)]^2\, d\omega. \qquad (2.35)$$

Выражение (2.35), получившее название равенства Парсеваля, показывает, что энергия сигнала может быть представлена в виде суммы бесконечно малых слагаемых $\frac{1}{\pi}[S(\omega)]^2\, d\omega$, соответствующих бесконечно малым участкам частотного спектра (рис. 2.23). Выражение $\frac{1}{\pi}[S(\omega)]^2$ представляет собой энергию, содержащуюся в спектральных составляющих сигнала, расположенных в полосе частот $d\omega$ в окрестности частоты ω.

Таким образом, квадрат модуля спектральной плотности характеризует распределение по спектру энергии сигнала.

Если задана энергия сигнала ΔW в определенной полосе частот $\Delta\omega$ в окрестности частоты ω_k (рис. 2.24), то модуль спектральной плотности в точке ω_k может быть найден из приближенного равенства

$$S(\omega_k) \approx \sqrt{\frac{\pi \Delta W}{\Delta\omega}}.$$

5. Практическая ширина спектра сигнала

Каждый реальный сигнал имеет конечную длительность и, следовательно, обладает бесконечным частотным спектром. Практически все каналы связи имеют ограниченную полосу пропускания. Следовательно, при передаче сигнала через реальный канал связи может быть передана лишь часть его частотного спектра. Поэтому приходится заботиться о том, чтобы обеспечить пропускание через канал связи наиболее существенной части спектра. В связи с этим введено понятие практической ширины спектра сигнала.

За практическую ширину спектра сигнала принимают диапазон частот, в пределах которого находится наиболее существенная часть спектра сигнала. Выбор практической ширины спектра сигнала определяется двумя критериями: энергетическим критерием и критерием допустимых искажений формы сигнала.

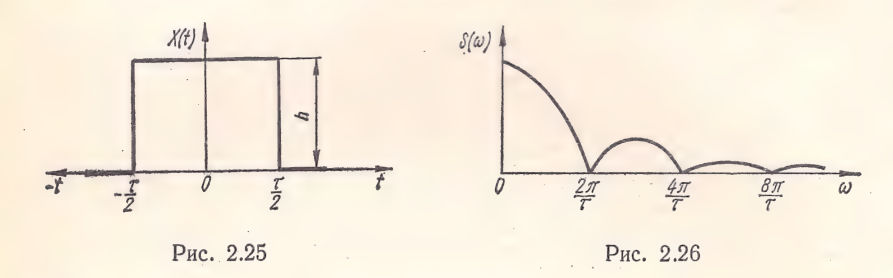


Рис. 2.25 Рис. 2.26

С энергетической точки зрения практическая ширина спектра периодического сигнала определяется как область частот, в пределах которой сосредоточена подавляющая часть всей энергии сигнала. Можно, например, показать (см. 2.3, пример 2.1), что для периодической последовательности импульсов прямоугольной формы длительностью $\tau = \frac{T}{2}$ достаточно практическую ширину спектра выбрать равной $5\omega_0 = 10\frac{\pi}{T} = 5\frac{\pi}{\tau}$. В этой области частот сосредоточено 96% всей мощности сигнала.

В качестве примера рассмотрим одиночный прямоугольный импульс длительностью τ и величиной h (рис. 2.25). Спектральная

плотность такого сигнала определяется выражением

$$S(j\omega)=\int_{-\infty}^{\infty}x(t)\exp\{-j\omega t\}\,dt=\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}}h\exp\{-j\omega t\}\,dt=\tau h\frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}} \tag{2.36}$$

и представлена на рис. 2.26.

В соответствии с формулой (2.35), энергия сигнала, сосредоточенная в полосе частот от 0 до ω_1,

$$W_1=\frac{1}{\pi}\int_0^{\omega_1}[S(\omega)]^2\,d\omega=\frac{\tau^2h^2}{\pi}\int_0^{\omega_1}\left[\frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}}\right]^2 d\omega.$$

Относительная величина энергии одиночного импульса, сосредоточенная в полосе частот от 0 до ω_1, выражается функцией

$$\lambda(\omega_1)=\frac{W_1}{W_0}=\frac{\tau}{\pi}\int_0^{\omega_1}\left[\frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}}\right]^2 d\omega, \tag{2.37}$$

где $W_0=\tau h^2$ — полная энергия одиночного прямоугольного импульса.

График функции $\lambda(\omega_1)$, называемый интегральной кривой распределения энергии сигнала в спектре, показан на рис. 2.27 [10, 11]. Как видно из этого рисунка, в полосе частот от 0 до $\omega_1=\frac{2\pi}{\tau}$ сосредоточено более 90% всей энергии сигнала и данный диапазон частот может быть взят в качестве практической ширины спектра одиночного прямоугольного импульса. При этом дальнейшее увеличение практической ширины спектра ведет к незначительному увеличению энергии в данном диапазоне частот, так как при $\omega_1>\frac{2\pi}{\tau}$ кривая $\lambda(\omega_1)$ довольно пологая.

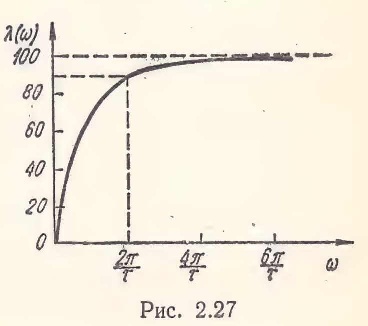


Рис. 2.27

Аналогично определяется практическая ширина спектра непериодических сигналов любой другой формы.

В тех случаях, когда важно сохранить требуемую форму сигнала, пользуются вторым критерием выбора практической ширины спектра сигнала.

Известно, что неискаженная передача формы сигнала будет обеспечена при условии, что комплексный коэффициент передачи канала связи имеет вид

$$\dot{K}(j\omega) = K \exp\{-j\omega T_0\}, \tag{2.38}$$

т. е. амплитудная и фазовая характеристики канала должны изменяться по закону

$$K(\omega) = K = \text{const}; \quad \varphi(\omega) = \omega T_0.$$

Передача сигналов по реальному каналу, обладающему ограниченной полосой пропускания, сопровождается искажением формы

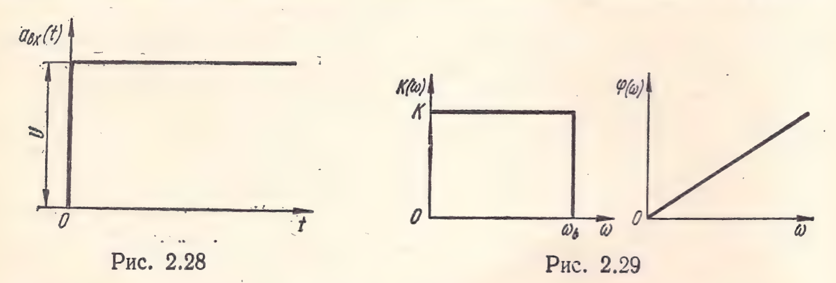


Рис. 2.28

Рис. 2.29

сигналов. На практике важное значение имеет обеспечение достаточно высокого значения крутизны или допустимого значения длительности фронта импульсов.

Для оценки влияния ширины полосы пропускания канала связи на искажение формы прямоугольных видеоимпульсов рассмотрим искажения ступенчатой функции $a_{вх} = U\,1(t)$, где $1(t)$ — единичная ступенчатая функция (рис. 2.28) при прохождении через канал связи, представляющий собой идеальный фильтр низких частот. Коэффициент передачи этого фильтра выражается зависимостью (2.38), при этом в диапазоне частот $0 \leqslant \omega \leqslant \omega_в$ модуль коэффициента передачи $K(\omega) = K = \text{const}$ и аргумент $\varphi(\omega) = \omega T_0$ [11, 12]; вне этого диапазона $K(\omega) = 0$ (рис. 2.29).

Спектральная плотность сигнала включения (см. пример 2.6) определяется выражением

$$S_{вх}(j\omega) = \frac{U}{j\omega}.$$

Тогда спектральная плотность выходного сигнала

$$S_{вых}(j\omega) = S_{вх}(j\omega)\dot{K}(\omega) = UK(\omega)\frac{\exp\{-j\varphi(\omega)\}}{j\omega}.$$

Использовав преобразование Фурье, получим выражение для выходного сигнала

$$a_{вых}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{вых}(j\omega)\exp\{j\omega t\}\,d\omega =$$

$$= \frac{UK(\omega)}{2\pi} \int\limits_{-\infty}^{\infty} \frac{\exp\{ j\omega (t - T_0) \}}{j\omega} d\omega = UK \left[\frac{1}{2} + \frac{1}{\pi} \int\limits_{0}^{\omega_в} \frac{\sin \omega t'}{\omega} d\omega \right], \quad (2.39)$$

где

$$t' = t - T_0.$$

Интегральный синус в выражении (2.39) является известной табличной функцией. Форма выходного сигнала, задержанного относительно входного на T_0, изображена на (рис. 2.30).

Фронт сигнала искажен и имеет сложный вид. В таких случаях рассматривают максимальное значение крутизны фронта, среднюю крутизну фронта и активную длительность фронта.

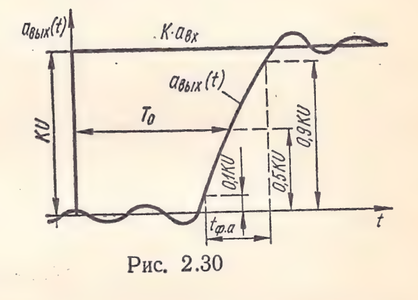


Рис. 2.30

Крутизна выходного сигнала

$$\frac{da_{вых}}{dt} = UK \int\limits_{0}^{\omega_в} \cos \omega t' d\omega = \frac{UK \sin \omega t'}{\pi t'}.$$

Наибольшее значение крутизны фронта достигается в момент $t = T_0 (t' = 0)$ и равно

$$\left(\frac{da_{вых}}{dt} \right)_{макс} = \lim_{t' \to 0} \frac{UK \sin \omega_в t'}{\pi t'} = 2UKf_в = 2U_{вых} f_в.$$

Активной длительностью фронта сигнала $t_{ф.а}$ считается промежуток времени, в течение которого сигнал изменяется от 0,1 до 0,9 установившегося значения. Средняя крутизна фронта на этом промежутке времени близка к максимальной крутизне. Тогда активная длительность фронта выходного сигнала

$$t_{ф.а} = \frac{0{,}9U_{вых} - 0{,}1U_{вых}}{\left| \frac{da_{вых}}{dt} \right|} \cong \frac{0{,}4}{f_в}. \quad (2.40)$$

Полученная форма выходного сигнала, вообще говоря, является нереальной, так как выходной сигнал появляется при $t < 0$, т. е. до начала поступления входного сигнала. Такой результат получен потому, что была принята физически неосуществимая модель канала связи, так как спектральная функция с ограниченным спектром может соответствовать только функции времени, существующей в интервале $-\infty < t < \infty$. Однако с точки зрения характеристики фронта импульсов полученный результат достаточно близко отражает истинную форму импульса.

На основании изложенного нетрудно определить искажение формы прямоугольного видеоимпульса длительности τ. Такой

импульс может быть представлен как наложение ступенчатых функций противоположного знака и смещенных на промежуток времени τ (рис. 2.31).

Аналитическое выражение для выходного сигнала может быть получено путем вычитания из функции (2.39) точно такой же функции, смещенной на промежуток времени τ

$$a_{\text{вых}}(t) = UK\left\{ \int_0^{\omega_\text{в}} \frac{\sin \omega t'}{\omega} \, d\omega - \int_0^{\omega_\text{в}} \frac{\sin[\omega(t' - \tau)]}{\omega} \, d\omega \right\}.$$

Очевидно, что формы переднего и заднего фронтов импульса будут одинаковы. Чтобы входной сигнал смог достичь наибольшего значения, длительность импульса τ должна быть примерно равна времени нарастания $t_{\text{ф. а}}$, т. е.

$$\tau \approx \frac{0{,}4}{f_\text{в}} \text{ или } f_\text{в} \approx \frac{0{,}4}{\tau}. \qquad (2.41)$$

Выражение (2.41) можно использовать для выбора практической ширины спектра прямоугольного импульса при заданной его длительности.

Практическая ширина спектра модулированных по амплитуде гармонических сигналов определяется законом изменения модули-

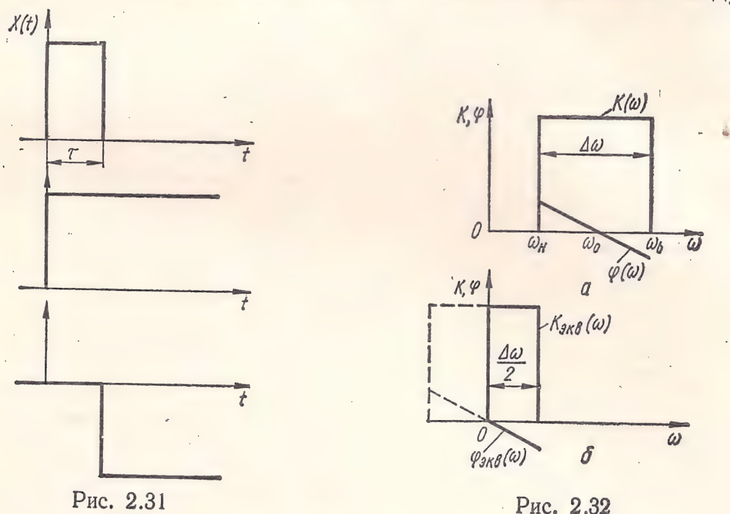


Рис. 2.31

Рис. 2.32

рующей функции и занимает определенную полосу частот в интервале $\omega_\text{н} \leqslant \omega \leqslant \omega_\text{в}$, посредине которого расположена несущая частота ω_0. В связи с этим для оценки влияния ограниченности полосы пропускания канала на искажение формы таких сигналов целесообразно канал связи рассматривать как идеальный полосовой фильтр с амплитудно-частотной и фазочастотной характеристиками, представленными на рис. 2.32, *а*.

Если фильтр имеет такие характеристики, а период колебания несущей частоты $t_0 = \frac{2\pi}{\omega_0} \ll \tau$, то для оценки искажения формы сигналов целесообразно воспользоваться теоремой об огибающей. Согласно этой теореме, спектральная плотность огибающей сигнала на выходе полосового фильтра с коэффициентом передачи $K(j\omega)$ равна произведению спектральной плотности огибающей входного сигнала и коэффициента передачи эквивалентного фильтра

$$K_{\text{экв}}(j\omega) = K[j(\omega_0 + \omega)].$$

Так как физически существующими являются только положительные частоты, то эквивалентный полосовой фильтр действует как фильтр низких частот, характеристики которого на рис. 2.32, *б* обозначены сплошными линиями. Следовательно, если, например, на вход фильтра поступает радиоимпульс прямоугольной формы, то огибающая выходного радиоимпульса будет иметь такую же форму, как и видеоимпульс прямоугольной формы такой же длительности после прохождения через фильтр низких частот с полосой пропускания $\Delta\omega = \frac{\omega_{\text{в}} - \omega_{\text{н}}}{2}$, при этом время нарастания выходного радиоимпульса

$$t_{\text{ф. а}} \approx \frac{0{,}8}{f_{\text{в}} - f_{\text{н}}} = \frac{1{,}6\pi}{\omega_{\text{в}} - \omega_{\text{н}}}. \tag{2.42}$$

Аналогичным способом могут быть оценены искажения каналами связи сигналов с другими видами модуляции.

2.3. Переход от преобразования Фурье к преобразованию Лапласа

При преобразованиях Лапласа временной функции $x(t)$ (оригиналу) соответствует некоторая определенная функция $x(p)$ (изображение) комплексного переменного $p = \sigma + j\omega$. Связь между изображением и его оригиналом выражается зависимостью

$$x(p) = \int_{-\infty}^{\infty} x(t) \exp\{-pt\}\, dt. \tag{2.43}$$

Обратное преобразование определяется выражением

$$x(t) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} x(p) \exp\{pt\}\, dp. \tag{2.44}$$

Преобразованию Лапласа удовлетворяют функции $x(t)$ при условии, что она непрерывна или имеет в любом конечном интервале

конечное число разрывов 1-го рода. Кроме того, интеграл (2.43) должен быть сходящимся; для этого достаточно, чтобы

$$|x(t)| \leqslant M \exp\{-\beta t\},$$

где M и β — любые положительные числа. Перечисленным требованиям удовлетворяют практически все реальные процессы.

Если сравнить выражение (2.43) для изображения $x(p)$ с выражением (2.28) для спектральной плотности, то можно увидеть, что спектральная плотность $S(j\omega)$ получается из изображения $x(p)$ заменой оператора $p = \sigma + j\omega$ на $j\omega$. При этом нужно иметь в виду, что замена $p = j\omega$ возможна лишь тогда, когда преобразование Лапласа не содержит полюсов, лежащих на мнимой оси и справа от нее. Следовательно, все соотношения спектрального метода можно получить из соответствующих соотношений операторного метода путем преобразования координат в плоскости комплексного переменного — оператора $p = \sigma + j\omega$.

Однако преобразование Фурье возможно только для ограниченного класса функций, удовлетворяющих требованию абсолютной сходимости интеграла $\int\limits_{-\infty}^{\infty} |x(t)|\, dt$.

Вместе с тем спектральный метод, несмотря на имеющиеся ограничения, нашел весьма широкое практическое применение. Хотя форма решения, полученная спектральным методом, оказывается иногда неудобной, так как решение имеет вид бесконечной суммы гармонических составляющих, синтез которой затруднителен или даже вовсе не приводит к известным функциям, метод весьма удобен тем, что позволяет свести рассмотрение сложных процессов к хорошо знакомым элементарным процессам — гармоническим колебаниям.

2.4. Примеры

Пример 2.1. Определить практическую ширину спектра периодической последовательности прямоугольных импульсов при ширине импульсов τ, равной $\frac{T}{2}$, если требуется учесть все гармонические составляющие сигнала, амплитуды которых более 0,2 от амплитуды первой гармоники.

Из выражения (2.22) видно, что число гармоник n, подлежащих учету, может быть получено из выражения

$$\frac{A_n}{A_1} = \frac{2h\,\pi}{n\pi\, 2h} = \frac{1}{n} = 0{,}2,$$

откуда $n = 5$.

Таким образом, практическая ширина спектра равна $5\omega_0 = \frac{5\pi}{\tau}$.

Определим, какая часть мощности сигнала сосредоточена в пределах практической ширины спектра.

В пределах практической ширины спектра размещаются первая, третья и пятая гармоники и постоянная составляющая. Следовательно, мощность сигнала, приходящаяся на практическую ширину спектра

$$P_n = \frac{h^2}{4} + \frac{1}{2}\left(\frac{2h}{\pi}\right)^2 + \frac{1}{2}\left(\frac{2h}{3\pi}\right)^2 + \frac{1}{2}\left(\frac{2h}{5\pi}\right)^2 \cong 0{,}48h^2.$$

Полная мощность сигнала

$$P_0 = h^2\left[\frac{1}{4} + \frac{2}{\pi^2}\left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots\right)\right] =$$

$$= h^2\left[\frac{1}{4} + \frac{2}{\pi^2}\frac{\pi^2}{8}\right] = 0{,}5h^2.$$

Таким образом, в пределах практической ширины спектра сосредоточено 96% всей мощности сигнала.

Пример 2.2. Найти спектр последовательности косинусоидальных импульсов (рис. 2.33).

Функция $x(t)$, описывающая данный сигнал, может быть представлена следующим образом:

$$x(t) = \begin{cases} h\cos\omega_0 t & \text{при } -\frac{\tau}{2} \leqslant t \leqslant \frac{\tau}{2}; \\ 0 & \text{при } -\frac{3\tau}{2} < t < -\frac{\tau}{2} \text{ и } \frac{\tau}{2} < t < \frac{3\tau}{2}; \end{cases}$$

$$\omega_0 = \frac{2\pi}{T}; \quad T = 2\tau.$$

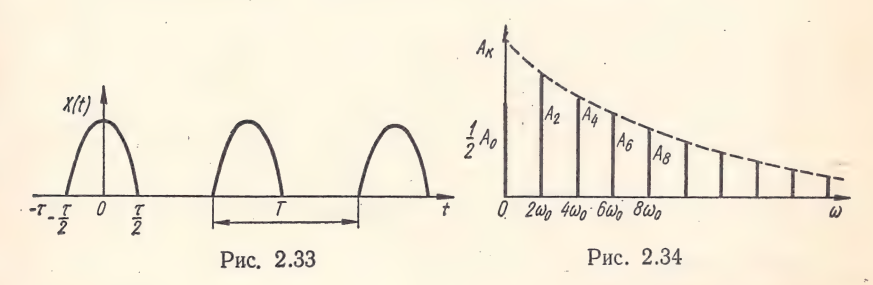


Рис. 2.33

Рис. 2.34

Комплексная амплитуда сигнала, согласно (2.6),

$$A_k = \frac{2}{T}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} h\cos\omega_0 t \exp\{-jk\omega_0 t\}\,dt =$$

$$= \frac{2h}{T}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \frac{\exp\{j\omega_0 t\} + \exp\{-j\omega_0 t\}}{2}\exp\{-jk\omega_0 t\}\,dt =$$

$$= \frac{h}{T}\left[\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \exp\{-j(k-1)\omega_0 t\}\,dt + \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \exp\{-j(k+1)\omega_0 t\}\,dt\right] =$$

$$= \frac{h}{T}\left[\frac{\exp\left\{j(k-1)\,\omega_0\frac{\tau}{2}\right\} - \exp\left\{-j(k-1)\,\omega_0\frac{\tau}{2}\right\}}{j(k-1)\,\omega_0} + \right.$$

$$\left. + \frac{\exp\left\{j(k+1)\,\omega_0\frac{\tau}{2}\right\} - \exp\left\{-j(k+1)\,\omega_0\frac{\tau}{2}\right\}}{j(k+1)\,\omega_0}\right] =$$

$$= \frac{h}{\pi(k-1)}\sin(k-1)\,\omega_0\frac{\tau}{2} + \frac{h}{\pi(k+1)}\sin(k+1)\,\omega_0\frac{\tau}{2} =$$

$$= \frac{-h}{\pi(k-1)}\cos\frac{k\omega_0\tau}{2} + \frac{h}{\pi(k+1)}\cos\frac{k\omega_0\tau}{2}.$$

Так как

$$\left|\cos\frac{k\omega_0\tau}{2}\right| = \left|\cos\frac{k\pi}{2}\right| = \begin{cases} 1 \text{ при } k \text{ четном,} \\ 0 \text{ при } k \text{ нечетном,} \end{cases}$$

то спектр сигнала содержит только четные гармоники, при этом комплексная амплитуда

$$\dot{A}_k = \frac{-2h}{\pi(k^2-1)}.$$

Модуль комплексной амплитуды

$$A_k = \frac{2h}{\pi(k^2-1)}.$$

График спектра амплитуд изображен на рис. 2.34.

Пример 2.3. Найти спектр одиночного прямоугольного импульса величиной h и длительностью τ (рис. 2.25) по известному спектру периодической последовательности таких же импульсов, следующих с частотой $\omega_0 = \frac{2\pi}{T}$.

Пользуясь связью между спектром непериодического сигнала и спектром соответствующего периодического сигнала, получим следующее выражение для спектральной плотности прямоугольного импульса

$$S(j\omega) = \frac{T}{2}\dot{A}(\omega),$$

где $\dot{A}(\omega)$ — огибающая спектра периодической последовательности импульсов.

Используя выражение (2.21), получим

$$S(j\omega) = h\tau\frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}}. \tag{2.45}$$

Отсюда модуль спектральной плотности

$$S(\omega) = h\tau \left| \frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}} \right|. \qquad (2.46)$$

График модуля спектральной плотности показан на рис. 2.26.

Пример 2.4. Найти спектр одиночного импульса высокочастотных колебаний (рис. 2.35).

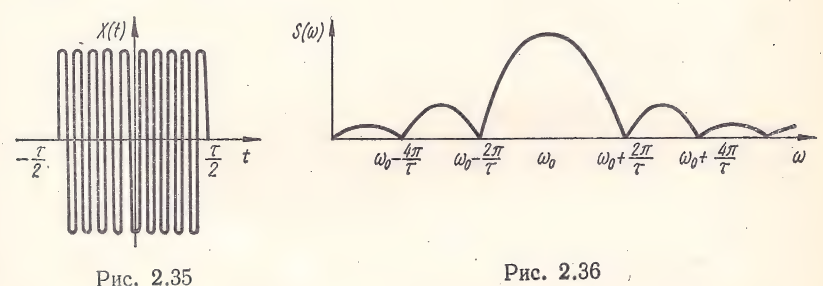


Рис. 2.35 Рис. 2.36

Функция $x(t)$, описывающая данный сигнал, может быть представлена в виде

$$x(t) = \begin{cases} h\cos\omega_0 t & \text{при } -\frac{\tau}{2} \leqslant t \leqslant \frac{\tau}{2}; \\ 0 & \text{при } \frac{\tau}{2} < t < -\frac{\tau}{2}. \end{cases}$$

Спектральная плотность такого сигнала равна

$$S(j\omega) = \int_{-\infty}^{\infty} x(t)\exp\{-j\omega t\}\,dt = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} h\cos\omega_0 t\exp\{-j\omega t\}\,dt =$$

$$= \frac{h}{2}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \exp\{j(\omega_0-\omega)t\}\,dt + \frac{h}{2}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \exp\{-j(\omega_0-\omega)t\}\,dt =$$

$$= \frac{h}{2}\left[\frac{\exp\left\{j(\omega_0-\omega)\frac{\tau}{2}\right\} - \exp\left\{-j(\omega_0-\omega)\frac{\tau}{2}\right\}}{j(\omega_0-\omega)} + \right.$$

$$\left. + \frac{\exp\left\{-j(\omega_0+\omega)\frac{\tau}{2}\right\} - \exp\left\{j(\omega_0+\omega)\frac{\tau}{2}\right\}}{j(\omega_0+\omega)}\right] =$$

$$= \frac{h}{(\omega_0 - \omega)} \sin(\omega_0 - \omega)\frac{\tau}{2} + \frac{h}{(\omega_0 + \omega)} \sin(\omega_0 + \omega)\frac{\tau}{2} =$$

$$= \frac{h\tau}{2} \frac{\sin\frac{(\omega_0 - \omega)\tau}{2}}{\frac{(\omega_0 - \omega)\tau}{2}} + \frac{h\tau}{2} \frac{\sin\frac{(\omega_0 + \omega)\tau}{2}}{\frac{(\omega_0 + \omega)\tau}{2}}. \qquad (2.47)$$

Из сравнения полученного выражения с выражением (2.46) для спектра одиночного импульса такой же длительности и величины h, но без высокочастотного заполнения, видно, что по отношению к спектру прямоугольного импульса спектр импульса высокочастотных колебаний смещен на величину несущей ω_0 и расширен в два раза за счет появления зеркального отображения спектра.

График модуля спектральной плотности одиночного импульса высокочастотных колебаний показан на рис. 2.36.

Пример 2.5. Определить спектр экспоненциального импульса (рис. 2.37).

Экспоненциальный импульс определяется функцией

$$x(t) = \begin{cases} h \exp\{-\beta t\} & \text{при } t \geqslant 0; \\ 0 & \text{при } t < 0. \end{cases} \qquad (2.48)$$

Спектральная плотность такого сигнала

$$S(j\omega) = \int_0^\infty h \exp\{-\beta t\} \exp\{-j\omega t\}\, dt = \frac{h}{\beta + j\omega}. \qquad (2.49)$$

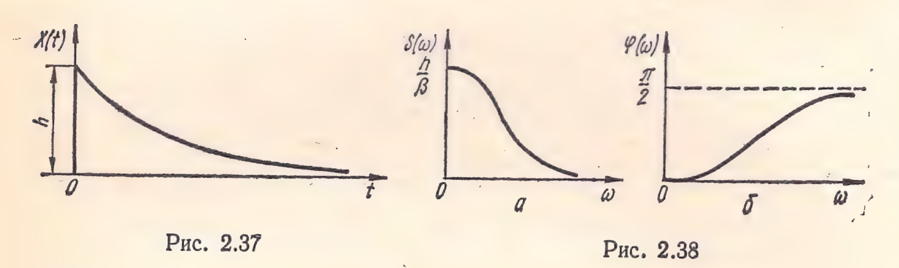


Рис. 2.37

Рис. 2.38

Модуль и фаза спектральной плотности определяются соответственно выражениями

$$S(\omega) = \frac{h}{\sqrt{\beta^2 + \omega^2}}; \quad \varphi(\omega) = \operatorname{arctg}\frac{\omega}{\beta}. \qquad (2.50)$$

На рис. 2.38, *а* и *б* приведены графики модуля и фазы спектральной плотности экспоненциального импульса.

Пример 2.6. Определить спектр сигнала включения (рис. 2.28). Сигнал включения определяется функцией

$$x(t) = h1(t) = \begin{cases} h & \text{при } t \geqslant 0; \\ 0 & \text{при } t < 0. \end{cases} \qquad (2.51)$$

Единичная функция $1(t)$ не удовлетворяет условию абсолютной интегрируемости и, следовательно, к ней нельзя применить преобразование Фурье. Однако ее можно рассматривать как образованную из импульса экспоненциальной формы (рис. 2.37) при неограниченном уменьшении его коэффициента затухания $\beta \to 0$.

В соответствии с этим спектральная плотность сигнала включения будет равна

$$S(j\omega) = \lim_{\beta \to 0} \frac{h}{\beta + j\omega} = \frac{1}{j\omega} = \frac{1}{\omega} \exp\left\{-j\frac{\pi}{2}\right\}, \quad (2.52)$$

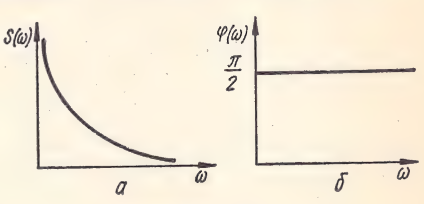


Рис. 2.39

откуда модуль и фаза спектральной плотности определяются выражениями

$$S(\omega) = \frac{1}{\omega}; \quad \varphi(\omega) = \frac{\pi}{2}. \quad (2.53)$$

Спектры амплитуд и фаз сигнала включения приведены на рис. 2.39, *а* и *б*.

Пример 2.7. Найти спектр дельта-функции.

Дельта-функцию можно трактовать как предел прямоугольного импульса длительности τ и амплитуды $\frac{1}{\tau}$, получаемый при $\tau \to 0$ (рис. 2.40).

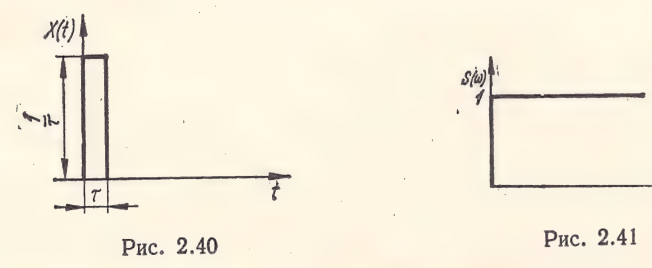


Рис. 2.40

Рис. 2.41

Тогда, исходя из (2.45), приняв амплитуду импульса равной $h = \frac{1}{\tau}$, получим

$$S_\delta(j\omega) = \lim_{\tau \to 0} \left[\frac{\sin\frac{\omega\tau}{2}}{\frac{\omega\tau}{2}}\right] = 1. \quad (2.54)$$

График спектральной плотности дельта-функции показан на рис. 2.41.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие детерминированные сигналы относятся к элементарным сигналам?
2. Какие способы представления моделей сигналов Вам известны?
3. В чем заключаются преимущества частотного метода представления сигналов?
4. При каких условиях периодическая функция может быть представлена рядом Фурье?
5. Что понимается под спектром амплитуд и спектром фаз?
6. Каковы характерные особенности спектра периодического сигнала?
7. При каких условиях непериодическая функция может быть представлена интегралом Фурье?
8. Каковы свойства спектральной плотности сигнала?
9. Каким образом можно получить спектр непериодического сигнала непосредственно из спектра соответствующего периодического сигнала?
10. Как можно энергетически истолковать спектр периодического и непериодического сигналов?
11. Что понимается под практической шириной спектра периодического и непериодического сигналов. Какие известны критерии выбора практической ширины спектра сигнала?
12. Сравните преобразование Фурье с преобразованием Лапласа.
13. Сравните ширины спектров амплитудно-модулированного, частотно-модулированного и фазомодулированного гармонических сигналов.
14. Чем определяется практическая ширина спектра периодической последовательности прямоугольных импульсов при условии, если длительность импульсов больше половины периода следования импульсов?
15. Как можно получить спектр импульса высокочастотных колебаний из спектра импульса, форма которого соответствует огибающей амплитуд импульса высокочастотных колебаний?
16. Как выглядит спектр экспоненциального импульса, сигнала включения и дельта-функции?

Глава III
МАТЕМАТИЧЕСКИЕ МОДЕЛИ СЛУЧАЙНЫХ СИГНАЛОВ

3.1. Случайные сигналы и их вероятностные характеристики

Как уже отмечалось, в реальных условиях все сигналы имеют случайный характер. Вследствие этого получателю заранее неизвестно, каким будет передаваемый сигнал.

Однако нельзя также утверждать, что приемная сторона не располагает абсолютно никакими предварительными (априорными) данными о сигналах. Во-первых, предварительно обычно известно все множество, весь ансамбль возможных сигналов. Во-вторых, имеются, как правило, сведения об ожидаемой вероятности тех или иных сигналов из общего ансамбля сигналов. Например, при передаче текста речи известно, какие буквы используются при передаче, какие буквы, сочетания букв и сочетания слов должны передаваться чаще и какие реже.

Таким образом, предварительные сведения, которыми мы располагаем о сигналах, носят статистический характер. Поэтому для иссле-

дования прохождения сигналов через информационные системы следует применять статистические методы.

Целесообразность применения статистических методов обусловлена еще и тем, что на сигнал воздействуют помехи, представляющие, как правило, неизвестную функцию времени. Основным содержанием задачи приема сигналов на фоне помех является наиболее полное извлечение информации из сигнала. Успешного решения этой задачи можно достичь только на основе использования статистических методов приема.

Наконец, целесообразность использования статистических методов вызвана еще и тем, что характеристики систем, по которым проходят сигналы, под влиянием разнообразных внешних и внутренних факторов могут изменяться случайным образом.

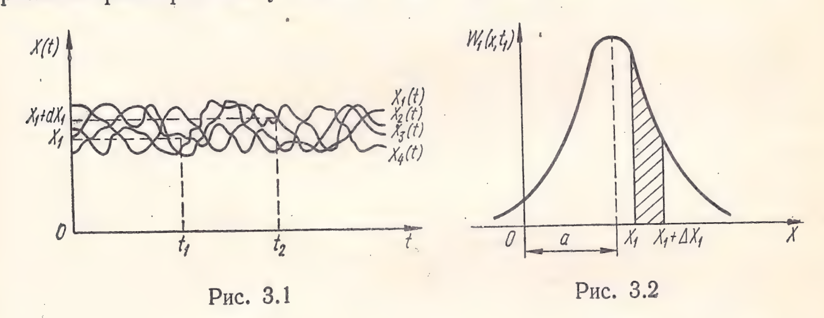


Рис. 3.1

Рис. 3.2

Процесс, описанием которого является случайная функция времени, называется случайным процессом. Иногда вместо термина «Случайный процесс» применяют в том же смысле термины «Стохастический процесс» или «Вероятностный процесс».

Конкретный вид, принимаемый случайным процессом в результате опыта, называется реализацией процесса. Отдельные наблюдения над случайным процессом, протекающим в однотипных системах при одинаковых условиях опыта, дадут различные реализации случайного процесса $x_1(t)$, $x_2(t)$, $\ldots$, $x_k(t)$ (рис. 3.1). Вид функции $x_k(t)$ случайным образом меняется от одного опыта к другому. Совокупность реализаций случайного процесса, полученных в результате опытов, называется ансамблем реализаций случайного процесса $x(t)$.

Величина k-й реализации случайного процесса в определенный момент времени (например, $t = t_1$) называется выборкой случайного процесса $x_k(t_1)$. Совокупность значений выборок в определенный момент времени (например, $t = t_1$) образует случайную величину $x(t_1)$.

Вероятность того, что в определенный момент времени $t = t_1$ величина x находится в интервале между x_1 и $x_1 + dx$,

$$P[x_1 \leqslant x(t_1) \leqslant x_1 + dx] = w_1(x,\ t_1)\,dx, \tag{3.1}$$

где $w(x,\ t_1)$ — одномерная плотность вероятности или одномерная функция распределения случайного процесса $x(t)$.

Плотность вероятности $w_1(x, t_1)$ есть в общем случае функция времени, так как зависит от t_1, является частной производной от интегральной функции распределения $F_1(x, t_1) = P(X_1 < x)$

$$w_1(x, t_1) = \frac{\partial F_1(x, t_1)}{\partial x}.$$

На рис. 3.2 приведен график наиболее часто встречающегося на практике нормального закона распределения плотности вероятности случайной величины x в определенный момент времени t_1.

Математическое описание этого закона имеет вид

$$w_1(x, t_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x(t_1) - a]^2}{2\sigma^2}\right\}, \tag{3.2}$$

где a и σ — математическое ожидание и среднеквадратическое отклонение случайной величины x.

Заштрихованная на рис. 3.2. площадь под кривой распределения равна вероятности нахождения x в интервале между x_1 и $x_1 + \Delta x_1$

$$P[x_1 \leqslant x(t_1) \leqslant x_1 + \Delta x_1] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x_1}^{x_1+\Delta x_1} \exp\left\{-\frac{[x(t_1) - a]^2}{2\sigma^2}\right\} dx. \tag{3.3}$$

При любом законе распределения плотности вероятности справедливо равенство

$$\int_{-\infty}^{\infty} w_1(x, x_1)\, dx = 1. \tag{3.4}$$

Одномерный закон распределения плотности вероятностей является простейшей статистической характеристикой случайного процесса. Он дает представление о процессе лишь в отдельные, фиксированные моменты времени; характеризует процесс статически и не дает представления о динамике его развития.

Для более полной характеристики случайного процесса необходимо знать связь между вероятными значениями случайной функции при двух произвольных моментах времени t_1 и t_2. Эта связь выражается через двумерную плотность вероятности и формулируется следующим образом: вероятность нахождения любой из функций $x_k(t)$, входящих в совокупность функций $x(t)$, в интервале $(x_1, x_1 + dx_1)$ в момент времени t_1 и в интервале $(x_2, x_2 + dx_2)$ в момент времени t_2

$$P[x_1 \leqslant x(t_1) \leqslant x_1 + dx_1; \quad x_2 \leqslant x(t_2) \leqslant x_2 + dx_2] = \\ = w_2(x_1, t_1; x_2, t_2)\, dx_1\, dx_2, \tag{3.5}$$

где $w_2(x_1, t_1, x_2, t_2)$ — двумерная плотность вероятности или двумерная функция распределения случайного процесса $x(t)$.

Дисперсия (или второй центральный момент одномерного закона распределения) — это математическое ожидание квадрата отклонения величин $x(t_1)$ от математического ожидания в определенный момент времени t_1, определяемое выражением

$$D[x(t_1)] = m_1\{[x(t_1) - m_1\{x(t_1)\}]^2\} = \\ = \int_{-\infty}^{\infty} \{x_1 - a[x(t_1)]\}^2 w_1(x, t_1)\,dx. \quad (3.8)$$

Дисперсия выражает меру разброса значений случайной величины $x(t_1)$ около математического ожидания, иными словами — «степень случайности» величины $x(t_1)$.

Корень квадратный из дисперсии принято называть среднеквадратическим отклонением случайной величины.

Аналогично можно найти среднее значение квадрата случайной величины $x(t_1)$

$$\sigma^2[x(t_1)] = m_1[x^2(t_1)] = \int_{-\infty}^{\infty} x_1^2 w_1(x, t_1)\,dx. \quad (3.9)$$

Положительный корень $\sigma[x(t_1)]$ из этой величины называется среднеквадратическим значением $x(t_1)$.

Из (3.8) можно установить следующую зависимость между дисперсией и среднеквадратическим значением случайной величины

$$D[x(t_1)] = \int_{-\infty}^{\infty} \{x^2 - 2xa[x(t_1)] + a^2[x(t_1)]\}\,w_1(x, t_1)\,dx = \\ = \int_{-\infty}^{\infty} x^2 w_1(x, t_1)\,dx - 2a[x(t_1)] \int_{-\infty}^{\infty} x w_1(x, t_1)\,dx + \\ + a^2[x(t_1)] \int_{-\infty}^{\infty} w_1(x, t_1)\,dx.$$

С учетом того, что $\int_{-\infty}^{\infty} w_1(x, t_1)\,dx = 1$, можно получить

$$D[x(t_1)] = \sigma^2[x(t_1)] - a^2[x(t_1)], \quad (3.10)$$

т. е. дисперсия равна разности квадратов среднеквадратического значения и математического ожидания.

При $a = 0$ дисперсии $D[x(t_1)]$ совпадают с квадратом среднеквадратического значения $\sigma^2[x(t_1)]$ случайной величины $x(t_1)$.

Математическое ожидание, дисперсия и среднеквадратическое значение случайной величины являются в общем случае функцией

времени. Они приближенно характеризуют поведение случайного процесса в отдельные моменты времени, но совершенно не затрагивают связь между значениями случайного процесса в различные моменты времени. Эту связь выражает корреляционная функция $K_{xx}(t_1, t_2)$, определяемая как среднее значение произведения значений случайной функции $x(t)$ в моменты времени t_1 и t_2

$$K_{xx}(t_1, t_2) = m_1[x(t_1)\, x(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 w_2(x_1, t_1;\ x_2, t_2)\, dx_1\, dx_2. \tag{3.11}$$

Связь между значениями двух случайных процессов $x(t_1)$ и $y(t_2)$ в моменты времени t_1 и t_2 соответственно выражает взаимная корреляционная функция $K_{xy}(t_1, t_2)$, определяемая как среднее значение произведения $x(t_1)\, y(t_2)$

$$K_{xy}(t_1, t_2) = m_1[x(t_1)\, y(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 y_2 w_2(x_1, t_1;\ y_2, t_2)\, dx_1\, dx_2. \tag{3.12}$$

Иногда рассматривают нормированную автокорреляционную r_{xx} и взаимную корреляционную r_{xy} функции (коэффициенты корреляции), определяемые выражениями

$$r_{xx}(t_1 t_2) = \frac{K_{xx}(t_1, t_2)}{\sigma[x(t_1)]\, \sigma[x(t_2)]}; \tag{3.13}$$

$$r_{xy}(t_1, t_2) = \frac{K_{xy}(t_1, t_2)}{\sigma[x(t_1)]\, \sigma[y(t_2)]}. \tag{3.14}$$

Нормированные корреляционные функции удобны тем, что они не превосходят единицы по абсолютной величине.

Математическое ожидание, дисперсия, квадрат среднеквадратического значения и корреляционная функция, определяемые выражениями (3.7), (3.8) и (3.11), получены путем осреднения по множеству реализаций случайного процесса для фиксированных моментов времени. Осредненные характеристики могут быть также получены путем обработки одной из реализаций случайного процесса на достаточно большом интервале времени.

Среднее по времени значение случайного процесса определяется выражением

$$\overline{x} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)\, dt, \tag{3.15}$$

где $x_i(t)$ — реализация случайного процесса $x(t)$; T — время наблюдения процесса.

По аналогии пользуются понятиями среднего по времени значения от функции $x^2(t)$, от квадрата разности $[x(t) - \overline{x}]^2$ и от произ-

ведения $x(t) \cdot x(t+\tau)$, определяемыми соответственно выражениями

$$\overline{x}^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i^2(t)\,dt; \qquad (3.16)$$

$$\overline{[x(t)-\overline{x}]^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x_i(t)-x]^2\,dt =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)\,dt - \lim_{T\to\infty} \frac{\overline{x}}{T} \int_{-T}^{T} x_i(t)\,dt + \overline{x}^2 = \overline{x}^2 - (\overline{x})^2; \quad (3.17)$$

$$\overline{x(t)\,x(t+\tau)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)\,x_i(t-\tau)\,dt. \qquad (3.18)$$

В общем случае для различных реализаций случайного процесса получаются различные значения среднего по времени от $x(t)$, $x^2(t)$, $[x(t)-x]$ и $x(t) \cdot x(t+\tau)$.

Если предположить, что $x(t)$ представляет изменение напряжения (или тока), то физически (3.15) равно мощности постоянной составляющей, рассеиваемой на сопротивлении в 1 Ом. В связи с этим считают, что среднее по времени $\overline{x}$ выражает мощность постоянной составляющей реализации случайного процесса $x(t_k)$.

По аналогии можно считать, что (3.16) выражает полную среднюю мощность, а (3.17) выражает среднюю мощность «случайной» составляющей процесса $x(t)$.

Если случайный сигнал является дискретным, то числовые характеристики определяются выражениями

$$\overline{x} = \sum_{k=1}^{n} P_k x_k;$$

$$\overline{x}^2 = \sum_{k=1}^{n} P_k x_k^2; \qquad (3.19)$$

$$[x(t)-\overline{x}]^2 = \sum_{k=1}^{n} P_k (x_k - \overline{x})^2;$$

$$\overline{x}_k \overline{x}_{k+e} = \sum_{k=1}^{n} P_{k,\,k+e}\, x_k x_{k+e},$$

где P_k — априорная вероятность случайной величины x_k; $P_{k,\,k+e}$ — совместная априорная вероятность величин x_k и $\overline{x}_{k+e}$; n — число значений случайной величины x.

3.3. Стационарные случайные процессы

В информационных системах очень часто встречаются случайные процессы, протекающие во времени приблизительно однородно. Эти процессы имеют вид непрерывных случайных колебаний около

некоторого среднего значения, причем ни среднее значение, ни характер этих колебаний не претерпевают существенных изменений во времени. Такие случайные процессы называются стационарными.

В качестве примеров стационарных случайных процессов можно привести шумы на выходе электронных устройств, случайные колебания в цепях питания и т. д.

В любой динамической системе случайный процесс начинается с так называемого «переходного» процесса и затем переходит в установившийся режим, который с некоторым приближением можно считать стационарным. Строго говоря, стационарные процессы бесконечны во времени, т. е. не имеют ни начала, ни конца. Таких процессов практически нет. Однако многие случайные процессы на определенных отрезках времени с определенным приближением можно считать стационарными.

Известно два понятия стационарности: стационарность в узком смысле и стационарность в широком смысле.

Под стационарными процессами в узком смысле понимаются случайные процессы, для которых функция распределения плотности вероятности $w_n(x_1, t_1;\ x_2, t_2, \ldots, x_n t_n)$ произвольного порядка n не меняется при любом сдвиге всей группы точек t_1, t_2, $\ldots$, t_n вдоль оси времени [14], т. е. для любых n и τ

$$w_n(x_1, t_1;\ x_2, t_2;\ \ldots;\ x_n, t_n) = w_n(x_1, t_1+\tau;\ x_2, t_2+\tau;\ \ldots;\ x_n, t_n+\tau). \tag{3.20}$$

Из приведенного определения следует, что для стационарных процессов:

а) одномерная функция распределения плотности вероятности не зависит от времени, т. е.

$$w_1(x, t_1) = w_1(x, t_1+\tau) = w_1(x); \tag{3.21}$$

б) двумерная функция распределения плотности вероятности зависит только от разности времени $t_2 - t_1 = \tau$, т. е.

$$w_2(x_1, t_1;\ x_2, t_2) = w_2(x_1, x_2;\ t_2 - t_1) = w_2(x_1, x_2, \tau); \tag{3.22}$$

в) трехмерная функция распределения плотности вероятности зависит только от двух разностей времен $t_2 - t_1 = \tau_1$ и $t_3 - t_1 = \tau_2$, т. е.

$$w_3(x_1, t_1;\ x_2, t_2;\ x_3, t_3) = w_3(x_1, x_2, x_3, t_2 - t_1;\ t_3 - t_1) = w_3(x_1, x_2, x_3, \tau_1, \tau_2) \text{ и т. д.} \tag{3.23}$$

Поскольку математическое ожидание и дисперсия выражаются через одномерную функцию распределения плотности вероятности, то на основании следствия (а) можно утверждать, что для стационарного процесса математическое ожидание и дисперсия не зависят от времени. Вследствие зависимости двумерной функции распределения только от разности времен $\tau = t_2 - t_1$, корреляционная функция стационарного процесса также зависит только от разности времен τ.

Таким образом, для стационарных процессов

$$\begin{aligned} a[x(t)] &= a(x); \\ D[x(t)] &= D(x); \\ K_{xx}(t_1, t_2) &= K_{xx}(\tau). \end{aligned} \tag{3.24}$$

На практике наиболее часто встречаются случайные процессы, для которых при выполнении условий (3.24) моменты высших порядков зависят от времени. Поэтому понятие стационарности оказалось целесообразным расширить, приняв за основу определения стационарности условия (3.24).

В связи с вышеизложенным введено понятие стационарности в широком смысле (или стационарности в смысле А. Я. Хинчина), согласно которому стационарными процессами являются такие случайные процессы, у которых математическое ожидание и дисперсия не зависят от времени, а корреляционная функция зависит только от разности времен $\tau = t_2 - t_1$.

Случайные процессы, стационарные в узком смысле, будут всегда стационарными в широком смысле, но не наоборот.

Стационарные случайные процессы по своей природе проще, чем нестационарные и описываются более простыми характеристиками. Ввиду того, что стационарные процессы встречаются на практике очень часто, получила широкое применение специальная теория стационарных случайных процессов. Раздел теории, посвященный изучению лишь тех свойств процессов, которые определяются моментами первых двух порядков, называется корреляционной теорией случайных процессов.

Так как свойства стационарного процесса во многом определяются свойствами корреляционной функции, то для изучения стационарного процесса нужно, в первую очередь, определить свойства корреляционной функции.

3.4. Свойства корреляционной функции стационарного случайного процесса

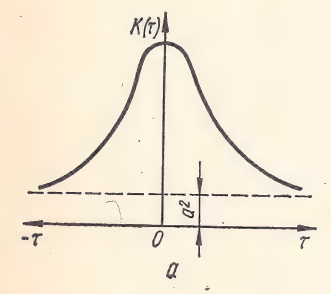
Определим, как ведет себя корреляционная функция при неограниченном увеличении разностного временного интервала $\tau = t_2 - t_1$.

По мере увеличения временного интервала τ зависимость между величинами $x(t)$ и $x(t+\tau)$ ослабевает. В пределе при $\tau \to \infty$ эти величины становятся независимыми. Тогда с учетом того, что математическое ожидание произведения случайных независимых величин равно произведению математических ожиданий сомножителей и что для стационарного процесса математическое ожидание не зависит от времени, получим

$$\begin{aligned} \lim_{\tau \to \infty} K_{xx}(\tau) &= \lim_{\tau \to \infty} m_1 [x(t)\, x(t+\tau)] = \\ &= \lim_{\tau \to \infty} [m_1 [x(t)]]\, m_1 [x(t+\tau)] = a^2(x). \end{aligned} \tag{3.25}$$

Корреляционная функция пульсаций стационарного случайного процесса равна

$$K_{xx}^{0}(\tau)=m_1[x_0(t)\,x_0(t+\tau)]=m_1\{[x(t)-a(x)][x(t+\tau)-$$
$$-a(x)\}=m_1[x(t)\,x(t+\tau)]-a(x)\,m_1[x(t+\tau)]-$$
$$-a(x)\,m_1[x(t)]+a^2(x)=K_{xx}(\tau)-a^2(x). \quad (3.30)$$



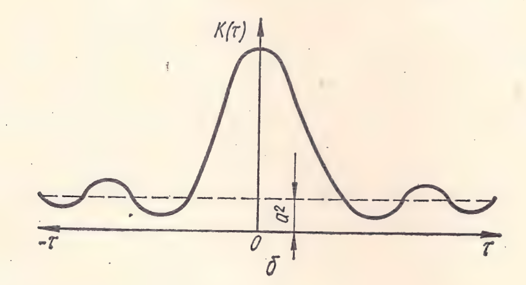


Рис. 3.4

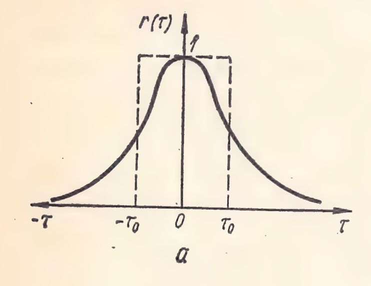
Из (3.30) следует, что математическое ожидание пульсаций равно нулю, а их дисперсия

$$D_0(x)=K_{xx}^{0}(0). \quad (3.31)$$

Отношение

$$r_x^0(\tau)=\frac{K_{xx}(\tau)}{D_0(x)} \quad (3.32)$$

называется нормированной корреляционной функцией (коэффициентом корреляции) пульсаций случайного процесса (или случайного процесса с нулевым средним). Типичные кривые нормированной корреляционной функции пульсаций показаны на рис. 3.5.



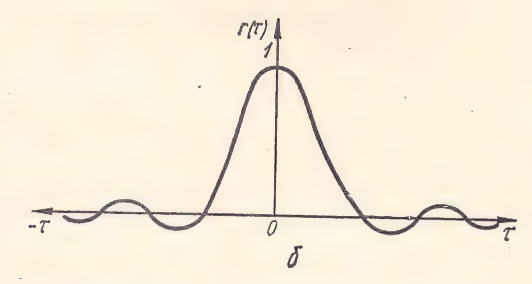


Рис. 3.5

Для чисто случайного стационарного процесса всегда можно указать значение интервала $\tau=\tau_0$, что при $\tau>\tau_0$ величины $x(t)$ и $x(t+\tau)$ можно считать практически независимыми, причем практическая независимость понимается в том смысле, что при $\tau>\tau_0$

абсолютная величина коэффициента корреляции остается меньше заданной величины δ

$$|r_x^0(\tau)| < \delta. \tag{3.33}$$

Величина δ обычно принимается равной 0,05. Интервал τ_0 называют временем корреляции случайного процесса. Время корреляции иногда определяется как половина ширины основания прямоугольника единичной высоты, площадь которого равна площади под графиком коэффициента корреляции (рис. 3.5, *а*), т. е.

$$\tau_0 = \frac{1}{2}\int_{-\infty}^{\infty} r_x^0(\tau)\,d\tau = \frac{1}{K_{xx}^0(0)}\int_0^{\infty} K_{xx}^0(\tau)\,d\tau. \tag{3.34}$$

3.5. Эргодичность стационарных процессов

Существует класс случайных процессов, обладающих важным для практических приложений свойством эргодичности.

Случайный процесс называется эргодическим, если усреднение по множеству с вероятностью, сколь угодно близкой к единице, равно усреднению по времени.

Следовательно, для эргодических процессов справедливы равенства

$$\begin{gathered} a(x) = m_1[x(t)] = \overline{x}; \\ D(x) = m_1[x(t) - m_1\{x(t)\}]^2 = [\overline{x(t) - \overline{x}}]; \\ \sigma^2(x) = m_1[x^2(t)] = \overline{x}^2; \\ K_{xx}(\tau) = m_1[x(t)\,x(t+\tau)] = \overline{x(t)\,x(t+\tau)}. \end{gathered} \tag{3.35}$$

Развернув выражения средних по множеству и по времени, получим

$$\begin{gathered} a(x) = \int_{-\infty}^{\infty} x w_1(x)\,dx = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_i(t)\,dt; \\ D(x) = \int_{-\infty}^{\infty}[x - a(x)]^2 w_1(x)\,dx = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[x_i(t) - x]^2\,dt; \\ \sigma^2(x) = \int_{-\infty}^{\infty} x^2 w_1(x)\,dx = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_i^2(t)\,w_1(x)\,dt; \\ K_{xx}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 w_2(x_1, x_2, \tau)\,dx_1\,dx_2 = \\ = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_i(t)\,x_i(t+\tau)\,dt. \end{gathered} \tag{3.36}$$

Эргодическое свойство случайного процесса имеет большое практическое значение. При исследовании таких процессов нет необходимости изучать большую совокупность реализаций, а достаточно лишь одной реализации, наблюдаемой в течение длительного промежутка времени. Например, статистические свойства флюктуационных шумов на выходе электронных усилителей можно изучать в течение достаточно продолжительного времени на одном усилителе, а затем результаты этого исследования распространить на все идентичные устройства.

3.6. Спектральная плотность стационарного случайного процесса

Как установлено в предыдущей главе, при изучении детерминированных сигналов весьма удобным оказался гармонический анализ. В связи с этим желательно использовать аппарат преобразований Фурье к случайным процессам. Однако непосредственно приложение классического гармонического анализа к случайным процессам невозможно в основном по двум причинам:

1) реализации случайного процесса $x_k(t)$ не удовлетворяют условию абсолютной интегрируемости

$$\int_{-\infty}^{\infty} |x_k(t)|\, dt < \infty;$$

2) для случайного процесса $x(t)$ частотный спектр (в классическом представлении) также является случайной функцией.

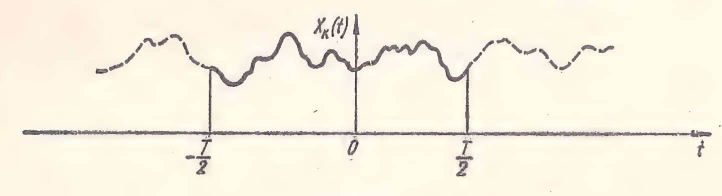


Рис. 3.6

Можно обобщить гармонический анализ, усредняя спектральные разложения, полученные для отдельных реализаций [14, 17].

Для этого введем новую функцию $x'_k(t)$, совпадающую на интервале $\left[-\frac{T}{2}; \frac{T}{2}\right]$ с реализацией $x_k(t)$ случайного процесса, а за пределами этого интервала равную нулю (рис. 3.6)

$$x'_k(t) = \begin{cases} x_k(t) & \text{при } t \leqslant \left|\frac{T}{2}\right|; \\ 0 & \text{при } t > \left|\frac{T}{2}\right|. \end{cases} \qquad (3.37)$$

то

$$m_1[G_k(\omega)] = \frac{1}{\pi T}\int_{-\frac{T}{2}}^{\frac{T}{2}} K_0(\tau)\exp\{-j\omega\tau\}\,d\tau \int_{-\frac{T}{2}}^{\frac{T}{2}} dt_2 =$$

$$= \frac{1}{\pi}\int_{-\frac{T}{2}}^{\frac{T}{2}} K_k(\tau)\exp\{-j\omega\tau\}\,d\tau.$$

Переходя к пределу при $T \to \infty$, окончательно получим

$$G(\omega) = \lim_{T\to\infty} G_k(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} K(\tau)\exp\{-j\omega\tau\}\,d\tau = \frac{2}{\pi}\int_{0}^{\infty} K(\tau)\cos\omega\tau\,d\tau. \quad (3.41)$$

Таким образом, спектральная плотность $G(\omega)$, являющаяся усредненной характеристикой совокупности реализаций случайного процесса, представляет собой прямое преобразование Фурье для корреляционной функции. Обратное преобразование Фурье имеет вид

$$K(\tau) = \frac{1}{2} - \int_{-\infty}^{\infty} G(\omega)\exp\{j\omega\tau\}\,d\omega = \int_{0}^{\infty} G(\omega)\cos\omega\tau\,d\omega. \quad (3.42)$$

Преобразования (3.41) и (3.42), связывающие функции $G(\omega)$ и $K(\tau)$, носят название преобразований Хинчина — Винера.

Если вместо круговой частоты ω ввести частоту f в герцах, то выражения (3.41) и (3.42) примут вид

$$K(\tau) = \int_{0}^{\infty} G(t)\cos 2\pi f\tau\,df; \quad (3.43)$$

$$G(f) = 4\int_{0}^{\infty} K(\tau)\cos 2\pi f\tau\,d\tau. \quad (3.44)$$

Так как для стационарных процессов усреднение по множеству может быть заменено усреднением по времени, то функция $G(\omega)$ может быть представлена в виде

$$G(\omega) = \lim_{T\to\infty}\frac{|S_k(j\omega)|^2}{\pi T}. \quad (3.45)$$

Таким образом, энергетический спектр стационарного случайного процесса может быть вычислен двумя путями:

а) непосредственным наблюдением одной реализации $x_k(t)$ и нахождением предела (3.45);

поскольку интеграл

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{j\omega\tau\}\, d\omega$$

является, как известно, обратным преобразованием Фурье от дельта-функции $\delta(\tau)$ (см. пример 3.13).

Таким образом, корреляционная функция белого шума выражается дельта-функцией (рис. 3.8, *б*)

$$K(\tau) = \begin{cases} \infty & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases}$$

Очевидно, что белый шум $x(t)$ характеризуется тем, что значения $x(t)$ в любые два (даже сколь угодно близкие) момента времени некоррелированы.

Такой случайный процесс иногда называют абсолютно случайным.

Необходимо подчеркнуть, что понятие «белый шум» основано на спектральном свойстве случайного процесса и совершенно не связано с законами распределения плотности вероятности. В частности, если белый шум имеет нормальный закон распределения, то его называют нормальным белым шумом.

Белый шум в точном смысле является идеализацией, никогда не встречающейся в реальных условиях, хотя бы потому, что достаточно близкие значения случайной функции практически всегда зависимы, а также и потому, что реальные процессы имеют конечную мощность, а для белого шума полная мощность процесса бесконечна. Однако подобная идеализация во многих важных практических случаях значительно упрощает математический анализ и не вносит сколько-нибудь существенных погрешностей.

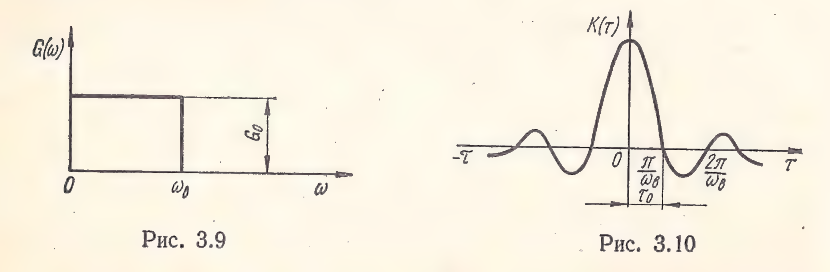


Рис. 3.9

Рис. 3.10

Спектры реальных процессов практически ограничены полосой частот $\Omega_э = \omega_в - \omega_н$, вследствие ограниченности полосы пропускания реальных систем.

Если белый шум пропустить через идеальный фильтр низких частот с граничными частотами $\omega_н = 0$, $\omega_в$, то на выходе получим шум с ограниченным спектром (рис. 3.9).

Корреляционная функция такого сигнала

$$K(\tau)=\int_0^{\omega_в} G_0 \cos\omega\tau\, d\omega = G_0\omega_в \frac{\sin\omega_в\tau}{\omega_в\tau} = P_0 \frac{\sin\omega_в\tau}{\omega_в\tau}, \qquad (3.48)$$

где $P_0 = G_0\omega_b$ — средняя мощность процесса.

График корреляционной функции (3.48) представлен на рис. 3.10.

Для такого процесса интервал корреляции τ_0 имеет конечную величину, которую можно определить, например, как интервал между точкой $\tau = 0$ и точкой, где $K(\tau)$ первый раз обращается в нуль, т. е. $\tau_0 = \frac{\pi}{\omega_в}$.

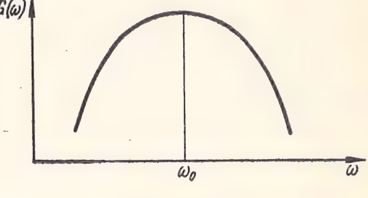


Рис. 3.11

Следовательно, ограничение спектра влечет за собой появление корреляции, причем по мере сокращения полосы частот $\Omega_э = \omega_в$ интервал корреляции увеличивается.

В случае, если для случайного процесса спектр непрерывен и сосредоточен около некоторой фиксированной частоты ω_0, причем выполняется условие

$$\frac{\Omega_э}{\omega_0} \ll 1, \qquad (3.49)$$

то такой процесс называется узкополосным.

Если узкополосный спектр обладает максимумом при ω_0 и симметричен относительно этой точки (рис. 3.11), то корреляционная функция процесса

$$K(\tau)=\int_0^{\infty} G(\omega)\cos\omega\tau\, d\omega = \int_{-\infty}^{\omega_0} G(\omega_0-\omega)\cos(\omega_0-\omega)\tau\, d\omega =$$

$$=\left[\int_{-\infty}^{\omega_0} G_0(\omega_0-\omega)\cos\omega\tau\, d\omega\right]\cos\omega_0\tau + \left[\int_{-\infty}^{\omega_0} G(\omega_0-\omega)\sin\omega\tau\, d\omega\right]\sin\omega_0\tau.$$

Поскольку по условию полоса спектра пренебрежимо мала по сравнению с частотой ω_0, то верхние пределы интегрирования без особой погрешности можно распространить до бесконечности.

Принимая во внимание, что второй интеграл, как интеграл от синуса в бесконечных пределах, равен нулю, получим

$$K(\tau) \cong \cos\omega_0\tau \int_{-\infty}^{\infty} G^*(\omega)\cos\omega\tau\, d\omega = b(\tau)\cos\omega_0\tau, \qquad (3.50)$$

где

$$C_0^*(\omega) = G(\omega_0-\omega);$$

$$b(\tau)=\int_{-\infty}^{\infty} G^*(\omega)\cos\omega\tau\, d\omega = 2\int_0^{\infty} G(\omega)\cos\omega\tau\, d\omega.$$

Из (3.50) видно, что корреляционная функция узкополосного процесса, спектр которого расположен симметрично около некоторой частоты ω_0, равна умноженной на $\cos \omega_0 \tau$ корреляционной функции $b(\tau)$, которая соответствует спектру $G^*(\omega)$, полученному из исходного смещением на величину ω_0 в область низких частот.

3.7. Эффективная ширина спектра случайного процесса

При анализе случайных процессов с неравномерным спектром (рис. 3.12) часто пользуются понятием эквивалентной или эффективной ширины спектра, определяемой выражением

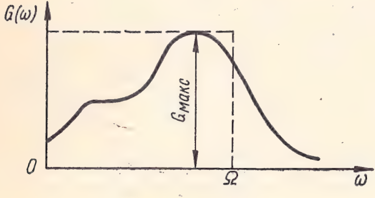


Рис. 3.12

$$\Omega_э = \frac{\int\limits_0^\infty G(\omega)\,d\omega}{G_{макс}(\omega)}, \qquad (3.51)$$

где $G_{макс}(\omega)$ — наибольшее значение функции спектральной плотности.

Средняя мощность процесса при этом будет равна

$$K(0) = G(0)\,\Omega_э. \qquad (3.52)$$

И так как интервал корреляции определяется с учетом (3.41) соотношением

$$\tau_0 = \frac{\int\limits_0^\infty |K(\tau)|\,d\tau}{K(0)} = \frac{\pi G(0)}{2K(0)},$$

то может быть установлена следующая связь между интервалом корреляции и эффективной шириной спектра процесса:

$$\tau_0 = \frac{\pi}{2\Omega_э} = \frac{1}{4F_э}. \qquad (3.53)$$

3.8. Кодирование сигналов. Модели сигналов при кодоимпульсной модуляции

При передаче дискретных сообщений должен передаваться условный номер сообщения. Это осуществляется обычно путем кодирования сигналов.

Под кодированием понимают преобразование дискретных сигналов в сложные сигналы (кодовые комбинации), представляющие собой набор некоторых элементарных сигналов.

Элементарными сигналами, называемыми также символами кода, в технических информационных системах обычно служат одиночные импульсы постоянного тока (видеоимпульсы) или переменного тока

(радиоимпульсы). Элементарным сигналом может служить пауза между импульсами или комбинации паузы и импульса и так далее.

Кодирование преследует две основные цели: повысить помехоустойчивость передачи и повысить эффективность передачи, т. е. обеспечить наилучшее согласование скорости передачи с пропускной способностью канала связи. В соответствии с этим кодовые сигналы должны строиться с учетом статистики источника информации, структуры помех в линии связи и параметров канала связи.

Кодирование обеспечивает изменение структуры сигналов и ни в коей мере не должно изменять количества информации, заключенной в первоначальном сообщении.

Элементарные сигналы (символы кода), составляющие кодовую комбинацию, должны различаться по какому-либо одному или нескольким параметрам, называемым часто кодовыми признаками. В качестве кодовых признаков применяются такие параметры, как величина, полярность, время (продолжительность или фаза импульсов), частота заполнения импульса и т. п.

Общее число символов, составляющих кодовую комбинацию, называется значностью, или длиной кода n. Количество значений кодовых признаков, используемых в кодовых комбинациях, называется основанием кода m.

На рис. 3.13, *а* приведен пример кодовой комбинации с длиной кода $n = 5$. В качестве импульсного признака здесь использована величина импульсов. Основание кода (число значений импульсного признака) $m = 3$. На рис. 3.13, *б* приведена кодовая комбинация с $n = 5$ и $m = 3$, в которой в качестве импульсного признака используется длительность импульсов.

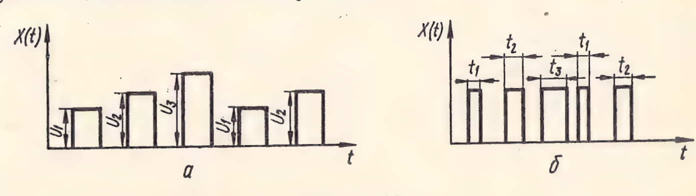


Рис. 3.13

По условиям построения кодовых комбинаций коды делятся на равномерные и неравномерные. В равномерных кодах все сообщения передаются кодовыми группами и с одинаковым числом элементов $n = \text{const}$.

При использовании неравномерных кодов разные сообщения могут передаваться кодовыми группами, содержащими не одинаковое число элементов ($n = \text{var}$). Равномерный код обладает большими возможностями с точки зрения обеспечения помехозащищенности передачи, так как потеря элементов или возникновение новых элементов в кодовых комбинациях с $n = \text{const}$ могут быть легко обнаружены. Неравномерные коды могут обеспечить наибольшую экономичность построения кодов и наибольшее быстродействие

передачи сообщений. Такие коды используются при так называемом статистическом кодировании.

Вместе с тем неравномерные коды менее помехозащищенные, чем равномерные. Потеря или возникновение новых элементов в комбинации в результате действия помех могут привести к созданию новой разрешенной комбинации, воспринимаемой на приемной стороне как истинной.

По числу различных символов (m) в кодовых комбинациях различают коды:

единичные, $m = 1$;

двоичные, $m = 2$;

коды с основанием $m > 2$.

В единичном коде используются одинаковые символы. Кодовые группы отличаются друг от друга лишь количеством символов (импульсов). Такие коды называют еще числоимпульсными. Иллюстрация числоимпульсного кода дается на рис. 3.14.

Единичный код отличается своей простотой. Однако, вследствие того что он неравномерен ($n = \text{var}$), помехозащищенность его низка. Кроме того, при передаче большого количества сообщений происходит изменение в широких пределах длины кода, что вызывает определенные неудобства.

Число	Код
1	
3	
6	
10	

Рис. 3.14

В связи с вышеизложенным единичный код практически не используется для передачи информации по каналу связи, а используется лишь при промежуточных преобразованиях сигналов на передающей и приемной сторонах.

Наибольшее распространение получили двоичные коды ($m = 2$). Это обусловлено следующим. Формирование кодовых сигналов и их дешифрация производятся с помощью релейных устройств, способных занимать ряд устойчивых состояний. Количество таких состояний определяется основанием кода. Очевидно, что простейшими релейными устройствами являются устройства с двумя состояниями. К такого типа устройствам принадлежит большинство электромагнитных реле, электронных и других типов бесконтактных реле. Кроме того, следует также учитывать простоту выполнения арифметических и логических операций при двоичном кодировании.

Коды с основанием $m > 2$ пока не нашли широкого применения в информационных системах.

По форме представления в канале передачи различают последовательные и параллельные коды. При последовательной форме элементарные сигналы, составляющие кодовую комбинацию, посылаются в канале передачи последовательно во времени. Как правило, они разделены между собой определенным временным интервалом.

При параллельной форме элементарные сигналы посылаются одновременно по нескольким электрическим цепям, число которых соответствует количеству элементов кода.

Параллельная форма представления кода, хотя и требует меньшего времени для передачи сообщений, используется для передачи информации по каналу связи редко. Практически параллельная форма кода при передаче информации по однопроводной линии связи используется лишь в тех случаях, когда в качестве импульсного признака применяется частота.

Параллельная форма представления кода часто используется при преобразовании аналоговых величин в код и обратных преобразованиях, в устройствах памяти, регистрации и при логической обработке информации.

По возможности обнаружения и исправления ошибок различают простые и корректирующие коды. В простых кодах ошибка в опознании хотя бы одного элемента кода приводит к неправильной регистрации передаваемого сообщения. Корректирующие коды строятся таким образом, что неправильное опознание одного или даже нескольких элементов кодовой комбинации не вызывает ошибочной регистрации передаваемого сообщения.

Наконец, по основным законам кодообразования коды можно разделить на два класса: числовые (арифметические) и комбинаторные (нечисловые). Первые основаны на известных системах счисления, вторые — на законах теории соединений.

В технических информационных системах используются преимущественно числовые коды. Система счисления, положенная в основу числового кода, обусловливает структуру кодовых комбинаций, основание системы счисления определяет основание m соответствующего кода.

Наиболее распространенным в настоящее время является позиционный принцип образования системы счисления. При таком принципе значение каждого символа зависит от его положения — позиции в ряду символов, составляющих число. Единица каждого следующего разряда больше единицы предыдущего разряда в m раз, где m — основание системы счисления. При этом любое n-разрядное число N с основанием m может быть представлено в виде суммы

$$N = \sum_{i=0}^{n} e_i m^i, \qquad (3.54)$$

где e_i — значение разрядного коэффициента i-го разряда.

Число возможных значений разрядного коэффициента каждого разряда равно основанию выбранной системы счисления.

Простые числовые коды содержат количество элементов, соответствующее числу разрядов числа N, выражаемого данным кодом. Каждый элемент кодовой комбинации может принимать m различных значений импульсного признака.

Максимально возможное число кодовых комбинаций определяется выражением

$$N_{\text{макс}} = m^n.$$

Практически наиболее широко используются двоичные коды, т. е. коды, базирующиеся на двоичной системе счисления. Математическая запись двоичного кода следующая:

$$N = \sum_{i=0}^{n-1} e_i 2^i.$$

Здесь разрядные коэффициенты e_i могут принимать два значения, которые принято условно обозначать 0 и 1.

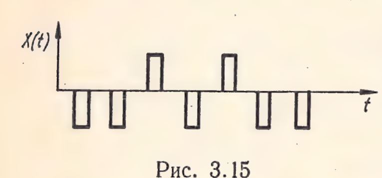


Рис. 3.15

Максимально возможное число кодовых комбинаций числового двоичного кода определяется выражением

$$N_{\text{макс}} = 2^n.$$

На рис. 3.15. показана двоичная кодовая комбинация длиной $n = 7$. В качестве импульсного признака здесь используется полярность. Импульсам положительной полярности соответствует 1 и импульсам отрицательной полярности — 0.

Кодовая комбинация содержит две единицы (два положительных импульса), но численное выражение (вес) каждой единицы определяется их номером (разрядом). Так как обычно нумерация разрядов идет справа, то номер младшего разряда является нулевым ($i = 0$) и численные выражения первого и второго (справа) импульсов соответственно будут

$$1 \cdot 2^2 = 4 \text{ и } 1 \cdot 2^4 = 16.$$

В целом код, представленный на рис. 3.15, выражает число

$$N = 0 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 0 \cdot 2^3 + 1 \cdot 2^4 + 0 \cdot 2^5 + 0 \cdot 2^6 = 20.$$

При передаче непрерывных сообщений в информационных системах весьма широкое применение получила кодоимпульсная модуляция (КИМ) сигналов. КИМ складывается из трех операций — дискретизации по времени, уровню и кодированию. Дискретизация по времени заключается в замене непрерывного во времени сигнала $x(t)$ дискретным сигналом $x_д(t)$, значения которого для дискретных моментов времени $t_0, t_1, t_2, \ldots, t_n$ совпадают соответственно с мгновенными значениями непрерывного сигнала (рис. 3.16, *а* и *б*). Такая операция называется квантованием сигналов по времени. Дискретизация по уровню (квантование по уровню) заключается в замене непрерывного множества значений сигнала $x_д(t)$ множеством дискретных значений. При этом шкала возможных значений сигнала разбивается на определенное количество интервалов и непрерывное значение сигнала заменяется ближайшим дискретным (рис. 3.16, *в*).

Полученные дискретные значения затем кодируются (обычно двоичным кодом).

КИМ обеспечивает существенное повышение помехоустойчивости передачи сообщений. Кроме того, дискретизация по времени позволяет использовать одни и те же устройства (каналы связи, устройства обработки информации и пр.) для большого числа различных сигналов.

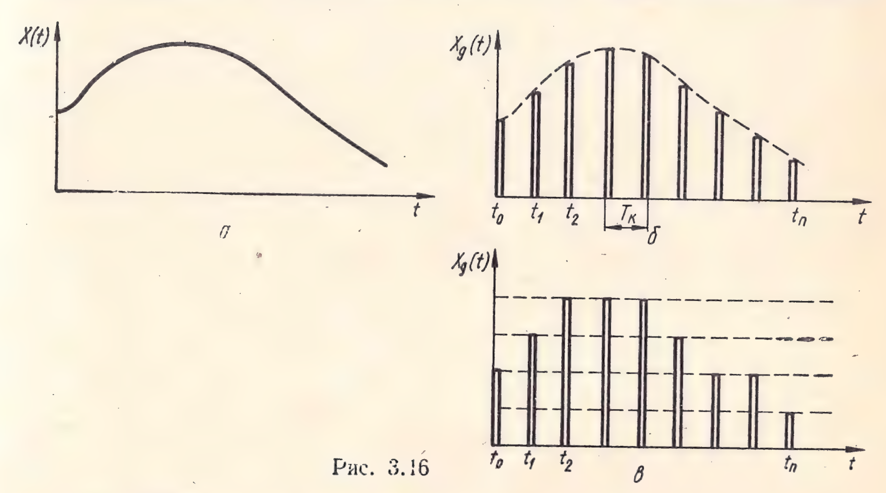


Рис. 3.16

При КИМ весьма важным является правильный выбор способа квантования сигналов по времени и уровню. В связи с этим рассмотрим некоторые вопросы теории квантования непрерывных функций по времени и уровню.

3.9. Квантование по времени

При квантовании по времени непрерывная по аргументу функция $x(t)$ преобразуется в функцию $x_д(t)$ дискретного аргумента. Такое преобразование может быть выполнено путем взятия отсчетов функции $x(t)$ в определенные дискретные моменты времени t_0, t_1, t_2, ..., t_n. В результате функция $x(t)$ заменяется совокупностью мгновенных значений $x(t_i)$ $i = 0, 1, 2, \ldots, n$.

Временной интервал $T_к = t_i - t_{i-1}$ между двумя соседними фиксированными моментами времени, в которых задается дискретная функция, называется интервалом временного квантования. Величина, обратная интервалу временного квантования

$$f_к = \frac{1}{T_к}, \qquad (3.55)$$

называется частотой квантования.

Частота квантования должна выбираться таким образом, чтобы по отсчетным значениям $x(t_i)$ можно было бы с заданной точностью получить исходную функцию.

Функцию $x(t)$, полученную в результате восстановления по отсчетам $x(t_i)$, принято называть воспроизводящей.

Известно несколько критериев выбора частоты квантования по времени. К таким критериям относится, в частности, частотный критерий В. А. Котельникова [4]. Данный критерий, который получил название теоремы Котельникова, основывается на следующей модели сигналов:

сигнал, представляет собой стационарный случайный процесс;

спектр сигнала сплошной и ограничен некоторой частотой, за пределами которой он тождественно равен нулю.

Теорема Котельникова формулируется следующим образом: если непрерывная функция $x(t)$ удовлетворяет условиям Дирихле (ограничена, кусочно-непрерывная и имеет конечное число экстремумов) и ее спектр ограничен некоторой частотой f_c, то она полностью определяется последовательностью своих значений в точках, отстающих на расстоянии $T_к = \frac{1}{2f_c}$ друг от друга.

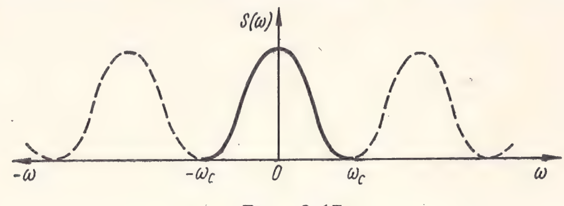


Рис. 3.17

Для доказательства теоремы рассмотрим выражения прямого и обратного преобразования Фурье непрерывной функции $x(t)$

$$S(j\omega) = \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t\}\, dt; \qquad (3.56)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) \exp\{j\omega t\}\, d\omega. \qquad (3.57)$$

В рассматриваемом частном случае функции с ограниченным спектром можно записать

$$x(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} S(j\omega) \exp\{j\omega t\}\, d\omega. \qquad (3.58)$$

Дополним функцию $S(j\omega)$ до периодической с периодом, равным $2\omega_c$ (рис. 3.17) и разложим ее в ряд Фурье

$$S(j\omega) = \sum_{-\infty}^{\infty} \dot{C}_к \exp\left\{j\pi k \frac{\omega}{\omega_c}\right\}, \qquad (3.59)$$

где

$$\dot{C}_к = \frac{1}{2\omega_c} \int\limits_{-\omega_c}^{\omega_c} S(j\omega) \exp\left\{-j\pi k \frac{\omega}{\omega_c}\right\} d\omega. \qquad (3.60)$$

Сравнивая выражения (3.58) и (3.60) замечаем, что они совпадают с точностью до постоянного множителя $\Delta t = \frac{\pi}{\omega_c}$, если принять $t = -k \cdot \Delta t$.

Следовательно,

$$\dot{C}_к = \frac{\pi}{\omega_c} x(-k\,\Delta t).$$

Подставив найденное выражение для $\dot{C}_к$ в (3.59), получаем

$$S(j\omega) = \sum_{k=-\infty}^{\infty} \frac{\pi}{\omega} x(-k\Delta t) \exp\left\{j\pi k \frac{\omega}{\omega_c}\right\}. \qquad (3.61)$$

После подстановки (3.61) в (3.58), замены знака при k (так как суммирование производится по всем положительным и отрицательным значениям k) и перестановки операций суммирования и интегрирования получим

$$x(t) = \frac{1}{2\omega_c} \sum_{k=-\infty}^{\infty} x(k\Delta t) \int\limits_{-\omega_c}^{\omega_c} \exp\{j\omega(t - k\Delta t\} d\omega. \qquad (3.62)$$

Вычислим интеграл

$$\int\limits_{-\omega_c}^{\omega_c} \exp\{j\omega(t - k\Delta t)\} d\omega = \int\limits_{-\omega_c}^{\omega_c} \cos\omega(t - k\Delta t)\, d\omega -$$

$$- j \int\limits_{-\omega_c}^{\omega_c} \sin\omega(t - k\Delta t)\, d\omega = \frac{2\sin\omega_c\ (t - k\Delta t)}{t - k\,\Delta t}, \qquad (3.63)$$

так как

$$\int\limits_{-\omega_c}^{\omega_c} \sin\omega(t - k\Delta t)\, d\omega = 0.$$

После подстановки (3.63) в (3.62) окончательно имеем

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\,\Delta t) \frac{\sin[\omega_c(t - k\Delta t)]}{\omega_c(t - k\Delta t)}. \qquad (3.64)$$

Полученное выражение представляет аналитически теорему Котельникова.

импульсы на всей оси времени от $-\infty$ до $+\infty$ или хотя бы достаточно большого количества импульсов до и после аппроксимируемого участка функции. Практически реализовать это трудно.

Кроме того, на практике приходится иметь дело с сигналами, ограниченными во времени и обладающими, следовательно, бесконечно широким спектром, что противоречит основному условию теоремы Котельникова.

Однако на практике никогда не требуется идеально точного воспроизведения передаваемого сигнала.

Поэтому с целью использования теоремы Котельникова для квантования сигналов реальный спектр сигнала, простирающийся от нуля до бесконечности, условно ограничивают некоторым диапазоном частот от нуля до ω_c, в котором сосредоточена основная часть энергии спектра (рис. 3.20). Энергия отсекаемой части спектра сигнала характеризует погрешность, возникающую за счет ограничений спектра сигнала частот ω_c.

Дисперсия приведенной погрешности, возникающей в результате ограничения частотного спектра сигнала,

$$D_{\omega_c} = \frac{\dfrac{1}{\pi T_c}\displaystyle\int_{\omega_c}^{\infty} |S(j\omega)|^2 d\omega}{(x_{\text{макс}} - x_{\text{мин}})^2}, \tag{3.65}$$

где $\dfrac{1}{\pi T_c}\displaystyle\int_{\omega_c}^{\infty} |S(j\omega)|^2 d\omega$ — средняя мощность отсекаемой части частотного спектра функции $x(t)$; T_c — длительность функции $x(t)$; $x_{\text{макс}}$, $x_{\text{мин}}$ — идеальные значения функции $x(t)$.

Для удобства расчетов целесообразней (3.65) представить в виде

$$D_{\omega_c} = \frac{Q_{\omega_c} P_T}{(x_{\text{макс}} - x_{\text{мин}})^2}, \tag{3.66}$$

где $P_T = \dfrac{1}{\pi T_c}\displaystyle\int_{0}^{\infty} |S(j\omega)|^2 d\omega = \int_{x_{\text{мин}}}^{x_{\text{макс}}} x^2 w(x)\,dx$ — средняя мощность сигнала; $Q_{\omega_c} = \dfrac{\displaystyle\int_{\omega_c}^{\infty} |S(j\omega)|^2 d\omega}{\displaystyle\int_{0}^{\infty} |S(j\omega)|^2 d\omega}$ — относительная площадь отсекаемого участка энергетического спектра функции $x(t)$.

Выражение (3.66) по заданной величине D_{ω_c} и при известных P_T, $x_{\text{макс}}$, $x_{\text{мин}}$ и спектре функции $x(t)$ позволяет определить частоту ω_c, ограничивающую спектр.

Дальнейшее развитие теория квантования по времени случайных процессов получила в работах Н. А. Железнова [19]. В основе модели сигналов, предложенной Железновым, лежат следующие допущения:

сигнал $x(t)$ представляет собой недетерминированный квазистационарный процесс;

длительность сигнала T_c конечная;

спектр сигнала сплошной и отличен от нуля на всей оси частот;

интервал корреляции сигнала $\tau_0 \ll T_c$;

мгновенная мощность сигнала ограничена и не может изменяться скачком.

Критерий Железнова формулируется следующим образом: квазистационарный сигнал с неограниченным спектром определяется со сколь угодной малой ошибкой последовательностью его мгновенных значений, следующих друг за другом через интервалы Δt, если $\Delta t \leqslant \tau_0$, а длительность сигнала $T_c \gg \tau_0$.

Ряд (3.64) применительно к критерию Железнова представится в виде

$$x(t) = \sum_{k=1}^{2f_c T_c} x(k\,\Delta t) \frac{\sin[\omega_c (t - k\Delta t)]}{\omega_c (t - k\Delta t)}, \qquad (3.67)$$

где

$$f_c = \frac{1}{2\Delta t}; \ \omega_c = 2\pi f_c.$$

Существенным преимуществом критерия Железнова является приближение модели сигнала к реальным условиям (неограниченность спектра и конечная длительность сигнала). Единственным ограничением является ограничение продолжительности функции корреляции величиной τ_0, определяемой выражением (3.34).

Трудности практической реализации методов Котельникова и Железнова обуславливают необходимость использования в ряде случаев других методов выбора требуемой частоты квантования сигналов по времени. В тех случаях, когда закон изменения функции с определенной достоверностью известен, более целесообразным является метод, основанный на замене непрерывной исходной функции аппроксимирующей функцией. В общем случае исходная функция аппроксимируется полиномом, кривая которого совпадает с кривой функции для заданных дискретных моментов времени.

Простейшим видом аппроксимации является аппроксимация полиномом первого порядка, при этом производится кусочно-линейная аппроксимация кривой функции, т. е. все точки кривой исходной функции, соответствующие отсчетным моментам времени, соединяются отрезками прямых (рис. 3.21). Выбор частоты квантования производится по критерию отклонения аппроксимирующей функции от исходной на каждом из интервалов дискретизации Δ_{t_i}. Чаще

других для этих целей применяются следующие критерии: критерий наибольшего отклонения, среднеквадратичный интегральный и вероятностный. В первом случае необходимая частота квантования выбирается из условия, чтобы предельные отклонения аппроксимирующей ломаной прямой от действительного значения функции (рис. 3.21) не превосходили бы заданного значения. Задача может быть решена с помощью интерполяционной формулы Ньютона, в соответствии с которой значение функции для любого момента времени внутри интервала $\Delta_t = t_{i+1} - t_i$ определяется выражением

$$x'(t) = x(t_i) + a_i(t - t_i),$$

где

$$a_i = \frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_1}.$$

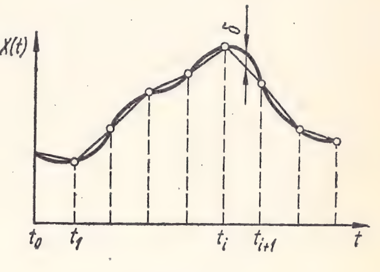


Рис. 3.21

Погрешность аппроксимации δ определяется остаточным членом $L(t)$ интерполяционной формулы

$$|\delta| = |L(t)| = x'(t) - x(t).$$

В рассматриваемом случае остаточный член выражается следующим образом:

$$L(t) = \frac{1}{2}\frac{d^2x(t)}{dt^2}(t - t_i)(t - t_{i+1}).$$

Очевидно, что максимальное значение погрешности аппроксимации

$$|\delta|_{\text{макс}} = |L(t)|_{\text{макс}} = \frac{1}{2}\left(\frac{d^2x(t)}{dt^2}\right)_{\text{макс}}\left(\frac{t_{i+1} - t_i}{2}\right)^2 = \frac{1}{8}\left(\frac{d^2x(t)}{dt^2}\right)_{\text{макс}} \cdot \Delta_t^2.$$

Следовательно, задаваясь допустимой погрешностью аппроксимации $\delta_{\text{макс}}$, можно определить интервал и частоту квантования

$$\Delta_t = \sqrt{\frac{8\delta_{\text{макс}}}{\left(\frac{d^2x(t)}{dt^2}\right)_{\text{макс}}}}; \quad f_{\text{к}} = \sqrt{\frac{\left(\frac{d^2x(t)}{dt^2}\right)_{\text{макс}}}{8\delta_{\text{макс}}}}.$$

Метод аппроксимации полиномом страдает определенной неточностью, обусловленной тем, что точный закон изменения функции $x(t)$ практически не известен, и поэтому невозможно точно определить максимальное значение второй производной функции.

Более высокую точность обеспечит аппроксимация полиномом, имеющим порядок выше первого. Кривая такой аппроксимирующей функции может состоять из отрезков дуг окружностей, отрезков парабол и пр.

3.10. Квантование по уровню

При квантовании по уровню непрерывное множество значений функции $x(t)$ заменяется множеством дискретных значений. Для этого в диапазоне непрерывных значений функции $x(t)$ выбирается конечное число дискретных значений этой функции (дискретных уровней) и в процессе квантования значение функции $x(t)$ в каждый момент времени заменяется ближайшим дискретным значением. В результате квантования образуется ступенчатая функция $x_д(t)$.

Квантование по уровню практически может осуществляться двумя способами. При первом способе мгновенное значение функции $x(t)$ заменяется меньшим дискретным значением (рис. 3.22, *а*).

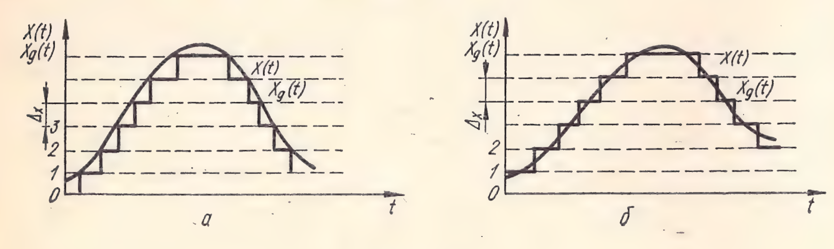


Рис. 3.22

При втором способе квантования мгновенное значение функций заменяется ближайшим меньшим или большим дискретным значением в зависимости от того, какое из этих значений ближе к мгновенному значению функции. В этом случае переход ступенчатой функции с одной ступени на другую происходит в те моменты, когда первоначальная непрерывная функция $x(t)$ пересекает середину между соответствующими соседними дискретными уровнями (рис. 3.22, *б*).

Расстояние между соседними дискретными уровнями называется интервалом или шагом квантования Δ_x.

Различают равномерное квантование по уровню, при котором шаг квантования постоянен, и неравномерное квантование по уровню, когда шаг квантования непостоянен. На практике преимущественное применение получило равномерное квантование в связи с простотой технической реализации.

Вследствие квантования функций по уровню появляются методические погрешности, так как действительное мгновенное значение функции заменяется дискретным значением. Эта погрешность, которая получила название погрешности квантования по уровню (или шума квантования), имеет случайный характер. Абсолютное ее значение в каждый момент времени определяется разностью между квантованным значением $x_д(t)$ и действительным мгновенным значением $x(t)$ функции

$$\delta_к(t) = x_д(t) - x(t).$$

На рис. 3.23, *а* и *б* показан характер изменения абсолютного значения погрешности квантования для способов квантования, представленных на рис. 3.22, *а* и 3.22, *б* соответственно. При квантовании методом замены действительного мгновенного значения функции $x(t)$ ближайшим меньшим дискретным значением абсолютная погрешность квантования будет всегда отрицательной и находиться в диапазоне $0 \div -\Delta_x$.

Закон распределения этой погрешности зависит от закона распределения $x(t)$.

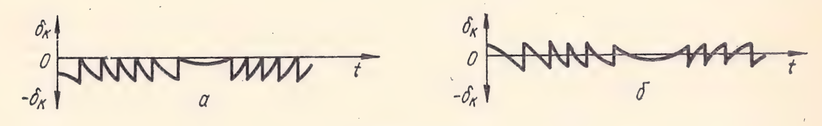


Рис. 3.23

Пусть функция $x(t)$ подчиняется определенному закону распределения $w(x)$ (рис. 3.24). Разобьем диапазон изменения функции $x(t)$ на интервалы величиной Δ_x. Пусть $\delta_к$ — случайное отклонение действительного значения функции $x(t)$ от ближайшего меньшего дискретного значения (погрешность квантования). Очевидно, что вероятность появления ошибки $\delta_к$ может быть определена как вероятность $p(\delta_к)$ попадания значения функции $x(t)$ в участок $\delta_к$ любого из интервалов квантования

$$p(\delta_к) = \sum_{i=0}^{n-1} \int_{i\Delta_x}^{i\Delta_x+\delta_к} w(x)\,dx.$$

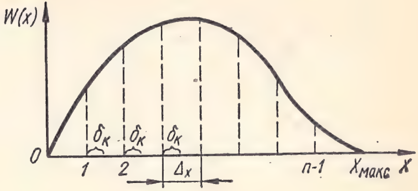


Рис. 3.24

Дифференцируя обе части данного равенства по $\delta_к$, найдем дифференциальный закон распределения погрешности квантования

$$\frac{dp(\delta_к)}{d\delta_к} = w(\delta_к) = \sum_{i=0}^{n-1} w(i\Delta_x + \delta_к).$$

Умножив обе части равенства на Δ_x, получим

$$w(\delta_к)\Delta_x = \sum_{i=0}^{n-1} w(i\Delta_x + \delta_к)\Delta_x. \quad (3.68)$$

Правую часть последнего равенства можно рассматривать как приближенное выражение площади, заключенной между осью x и кривой $w(x)$. Это приближение будет тем точнее, чем больше число участков квантования. Следовательно, в пределе при $n \to \infty$

или $\Delta_x \to 0$ выражение (3.68) приводится к виду

$$w(\delta_к)\,\Delta_x = \int\limits_0^{x_{макс}} w(x)\,dx. \qquad (3.69)$$

Так как правая часть выражает вероятность нахождения функции $x(t)$ в пределах от 0 до $x_{макс}$ и, следовательно, равна единице, то плотность вероятностей распределения погрешностей квантования будет с достаточным приближением

$$w(\delta_к) = \frac{1}{\Delta_x}. \qquad (3.70)$$

Таким образом, можно полагать, что при достаточно большом числе уровней квантования погрешность квантования подчиняется

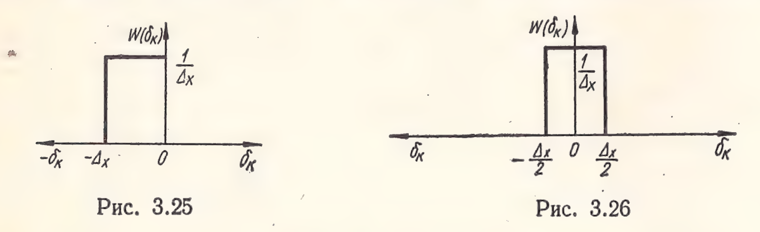


Рис. 3.25

Рис. 3.26

закону распределения равной вероятности, представленному на рис. 3.25.

В соответствии с рис. 3.25, можно записать, что

$$w(\delta_к) = \begin{cases} \dfrac{1}{\Delta_x} \text{ при } -\Delta_x \leqslant \delta_к \leqslant 0; \\ 0 \text{ при } \delta_к > 0 \text{ и } \delta_к < -\Delta_x. \end{cases}$$

Математическое ожидание погрешности квантования

$$a(\delta_к) = \int\limits_{-\Delta_x}^{0} \delta_к w(\delta_к)\,d\delta_к = \frac{1}{\Delta_x}\int\limits_{-\Delta_x}^{0} \delta_к\,d\delta_к = -\frac{\Delta_x}{2}. \qquad (3.71)$$

Дисперсия погрешности квантования

$$D(\delta_к) = \int\limits_{-\Delta_x}^{0} [\delta_к - a(\delta_к)]^2 w(\delta_к)\,d\delta_к = \int\limits_{-\Delta_x}^{0} \left[\delta_к + \frac{\Delta_x}{2}\right]^2 \frac{1}{\Delta_x}\,d\delta_к = \frac{\Delta_x^2}{12}. \qquad (3.72)$$

При квантовании методом замены действительного мгновенного значения функции ближайшим меньшим или большим дискретным значением погрешность квантования также подчиняется закону распределения равной вероятности, но изменяется в пределах от $-\frac{\Delta_x}{2}$ до $+\frac{\Delta_x}{2}$ (рис. 3.26).

В данном случае справедливо, что

$$w(\delta_к) = \begin{cases} \frac{1}{\Delta_x} \text{ при } -\frac{\Delta_x}{2} \leqslant \delta_к \leqslant \frac{\Delta_x}{2}; \\ 0 \text{ при } \delta_к > \frac{\Delta_x}{2} \text{ или } \delta_к < -\frac{\Delta_x}{2}. \end{cases}$$

Математическое ожидание погрешности квантования

$$a(\delta_к) = 0.$$

Дисперсия погрешности

$$D(\delta_к) = \int_{-\frac{\Delta_x}{2}}^{\frac{\Delta_x}{2}} \delta_к^2 w(\delta_к)\, d\delta_к = \frac{\Delta_x^2}{12}.$$

Практически в процессе квантования могут иметь место случаи, когда дискретные уровни не фиксированы относительно нулевого уровня функции $x(t)$ (рис. 3.27).

Так как при отсутствии фиксации дискретных уровней квантование производится практически путем замены мгновенного значения функции ближайшим меньшим дискретным уровнем и нижний дискретный уровень принимается за первый, то будут иметь место две

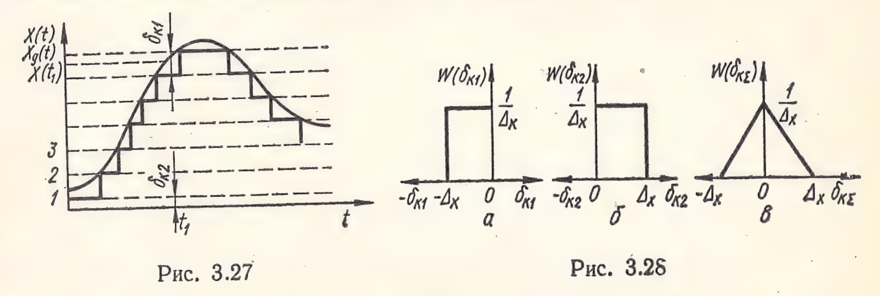


Рис. 3.27

Рис. 3.28

погрешности квантования: $\delta_{к_1}$, изменяющаяся в пределах от 0 до $-\Delta_x$, и $\delta_{к_2}$, изменяющаяся в пределах от 0 до $+\Delta_x$. Эти погрешности независимы и подчиняются законам распределения равной вероятности (рис. 3.28, *а* и *б*).

Суммарная погрешность квантования

$$\delta_{к_\Sigma} = \delta_{к_1} + \delta_{к_2}$$

изменяется в пределах от $-\Delta_x$ до $+\Delta_x$. Закон распределения суммарной погрешности является композицией законов распределения погрешностей $\delta_{к_1}$ и $\delta_{к_2}$ и выражается распределением Симпсона (рис. 3.28, *в*).

Аналитически данное распределение выражается следующим образом:

$$w(\delta_{\text{к}_\Sigma}) = \begin{cases} \dfrac{1+\dfrac{\delta_{\text{к}_\Sigma}}{\Delta_x}}{\Delta_x} & \text{при } -\Delta_x \leqslant \delta_{\text{к}_\Sigma} < 0; \\ \dfrac{1-\dfrac{\delta_{\text{к}_\Sigma}}{\Delta_x}}{\Delta_x} & \text{при } 0 \leqslant \delta_{\text{к}_\Sigma} \leqslant \Delta_x; \\ 0 & \text{при } \delta_{\text{к}_\Sigma} > \Delta_x \text{ и } \delta_{\text{к}_\Sigma} < -\Delta_x. \end{cases}$$

Математическое ожидание суммарной погрешности квантования

$$a(\delta_{\text{к}_\Sigma}) = 0;$$

дисперсия погрешности

$$D(\delta_{\text{к}_\Sigma}) = \int_{-\Delta_x}^{0} \delta_{\text{к}_\Sigma}^2 \frac{\left(1+\dfrac{\delta_{\text{к}_\Sigma}}{\Delta_x}\right)}{\Delta_x} d\delta_{\text{к}_\Sigma} + \int_{0}^{\Delta_x} \frac{1-\dfrac{\delta_{\text{к}_\Sigma}}{\Delta_x}}{\Delta_x} d\delta_{\text{к}_\Sigma} = \frac{\Delta_x^2}{6}.$$

Таким образом, в случае отсутствия фиксации дискретных уровней относительно начального уровня квантуемой функции среднеквадратическая погрешность квантования увеличивается в $\sqrt{2}$ раз.

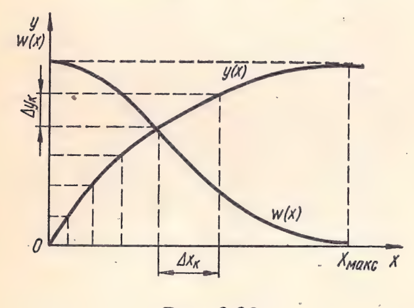


Рис. 3.29

В некоторых случаях, когда вероятность распределения значения функции по шкале уровней неодинакова, вводится неравномерное квантование по уровню с целью уменьшения среднеквадратической погрешности квантования. При таком квантовании значения функции x, вероятность возникновения которых велика, передаются с меньшей погрешностью квантования и маловероятные значения с большей погрешностью квантования [20].

Для реализации такого квантования первичная функция подвергается нелинейному преобразованию $y(x)$. Сущность нелинейного преобразования иллюстрируется рис. 3.29, где $w(x)$ — кривая распределения плотности вероятностей значения функции $x(t)$.

Величина y квантуется равномерно, при этом x квантуется неравномерно. Тогда между погрешностями квантования функций y и x будет иметь место следующая связь:

$$\delta y_{\text{к}} = \delta x_{\text{к}} = \frac{dy}{dx}.$$

Для дисперсии погрешностей получим

$$D_{y_к} = D_{x_к}\left(\frac{dy}{dx}\right)^2.$$

Тогда усредненная по параметру дисперсия погрешности квантования

$$\bar{D}_{x_к} = \int\limits_{x_{мин}}^{x_{макс}} w(x)\frac{D_{y_к}}{\left(\frac{dy}{dx}\right)^2}dx = D_{y_к}\int\limits_{x_{мин}}^{x_{макс}} \frac{w(x)}{\left(\frac{dy}{dx}\right)^2}dx.$$

Здесь множитель $D_{y_к}$ вынесен за знак интеграла, так как квантование функции y равномерное и дисперсия погрешности квантования не зависит от y и, следовательно, от x.

Выигрыш в погрешности квантования оценивается отношением

$$B = \frac{D_{y_к}}{D_{x_к}} = \frac{1}{\int\limits_{x_{мин}}^{x_{макс}} \frac{w(x)}{\left(\frac{dy}{dx}\right)^2}dx},$$

где $D_{x_к}$ — дисперсия погрешности равномерного квантования, равная $D_{y_к}$.

Для получения максимального выигрыша в погрешности квантования необходимо отыскать такую функцию $\frac{dy}{dx}$, которая минимизировала бы интеграл

$$\int\limits_{x_{мин}}^{x_{макс}} \frac{w(x)}{\left(\frac{dy}{dx}\right)^2}dx.$$

Решение подобной задачи приводится в [21].

3.11. Примеры

Пример 3.1. Найти математическое ожидание числа появлений событий A в одном испытании, если вероятность события A равна p.

Решение. Случайная величина x — число появлений событий A в одном испытании — может принимать только два значения: $x_1 = 1$ (событие A наступило) с вероятностью p и $x_2 = 0$ (событие A не наступило) с вероятностью $q = 1 - p$. Искомое математическое ожидание равно

$$a(x) = p\cdot 1 + q\cdot 0 = p.$$

Итак, математическое ожидание числа появлений события в одном испытании равно вероятности этого события.

Пример 3.2. Производится три выстрела с вероятностями попадания в цель, равными $p_1 = 0{,}4$; $p_2 = 0{,}3$ и $p_3 = 0{,}6$. Найти математическое ожидание общего числа попаданий.

Решение. Число попаданий при первом выстреле есть случайная величина x_1, которая может принимать только два значения: 1 (попадание) с вероятностью $p_1 = 0{,}4$ и 0 (промах) с вероятностью $q = 1 - 0{,}4 = 0{,}6$.

Математическое ожидание числа попаданий при первом выстреле равно вероятности попадания (см. пример 3.1), т. е. $a(x_1) = 0{,}4$.

Аналогично найдем математические ожидания числа попаданий при втором и третьем выстрелах

$$a(x_2) = 0{,}3; \quad a(x_3) = 0{,}6.$$

Общее число попаданий есть также случайная величина, состоящая из суммы попаданий в каждом из трех выстрелов

$$x = x_1 + x_2 + x_3.$$

Искомое математическое ожидание равно

$$a(x) = a(x_1) + a(x_2) + a(x_3) = 0{,}4 + 0{,}3 + 0{,}6 = 1{,}3.$$

Пример 3.3. Найти дисперсию случайной величины x, которая задана следующим законом распределения:

Таблица 1

x	1	2	5
p	0,3	0,5	0,2

Решение. В соответствии с (3.19, *a*) дисперсия дискретной случайной величины

$$D(x) = \sum_{k=1}^{3} p_к [x_к - a(x)]^2.$$

Математическое ожидание

$$a(x) = 0{,}3 \cdot 1 + 0{,}5 \cdot 2 + 0{,}2 \cdot 5 = 2{,}3.$$

Тогда искомая дисперсия

$$D(x) = 0{,}3\,(1 - 2{,}3)^2 + 0{,}5 \cdot (2 - 2{,}3)^2 + 0{,}2\,(5 - 2{,}3)^2 = 2{,}01.$$

Пример 3.4. Производится 10 независимых испытаний, в каждом из которых вероятность появления события A равна $p = 0{,}6$. Найти дисперсию числа появлений события A при испытаниях.

Решение. Число x появлений события A в $n = 10$ независимым испытаниям складывается из появлений события A в отдельных испытаниях, т. е.

$$x = x_1 + x_2 + x_3 + \cdots + x_{10}.$$

Величины $x_1, x_2, x_3, \ldots, x_{10}$ взаимно независимы, поэтому дисперсия суммы случайных величин равна сумме дисперсий этих величин

$$D(x) = D(x_1) + D(x_2) + D(x_3) + \cdots + D(x_n).$$

На основании (3.10) можно написать следующее выражение для дисперсии величин x_1:

$$D(x_1) = \sigma^2(x_1) - a^2(x_1).$$

Математическое ожидание числа появлений события в одном испытании равно вероятности события (см. пример 3.1), поэтому

$$a(x_1) = p.$$

Величина $\sigma^2(x_1)$ представляет собой математическое ожидание случайной величины x_1^2.

Так как x_1^2 может принимать только два значения, а именно 1^2 с вероятностью p и 0^2 с вероятностью q, то

$$\sigma^2(x_1) = m_1(x_1^2) = p \cdot 1^2 + q \cdot 0 = p.$$

Тогда

$$D(x_1) = p - p^2 = p(1 - p) = pq.$$

Очевидно для случайных величин $x_2, x_3, \ldots, x_{10}$ аналогично будем иметь

$$D(x_2) = pq; \quad D(x_3) = pq, \ \ldots; \quad D(x_{10}) = pq.$$

Следовательно, искомая дисперсия

$$D(x) = npq = 10 \cdot 0{,}6 \cdot 0{,}4 = 2{,}4.$$

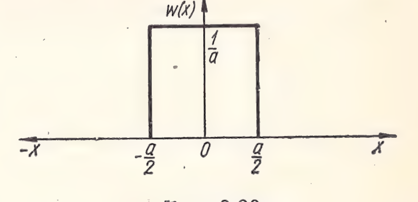


Рис. 3.30

Пример 3.5. В измерительном приборе расстояние между соседними метками шкалы постоянно и равно a. При округлении отсчета до ближайшего целого деления погрешность округления по абсолютной величине не превышает половины расстояния между соседними метками шкалы.

Найти плотность распределения вероятности, математическое ожидание и дисперсию погрешности округления.

Решение. Погрешность при округлении отсчета можно рассматривать как случайную величину x, которая может принимать с равной вероятностью любые значения в пределах от $-\frac{a}{2}$ до $+\frac{a}{2}$. Следовательно, плотность распределения случайной величины x постоянна в пределах от $-\frac{a}{2}$ до $+\frac{a}{2}$ и равна нулю за этими пределами (рис. 3.30).

Так как должно быть справедливо равенство

$$\int_{-\frac{a}{2}}^{\frac{a}{2}} w_1(x)\,dx = w_1(x)\int_{-\frac{a}{2}}^{\frac{a}{2}} dx = 1,$$

то

$$w_1(x) = \frac{1}{a}.$$

Аналитически закон равномерного распределения можно записать

$$w_1(x) = \begin{cases} 0 & \text{при } x < -\frac{a}{2}; \\ \frac{1}{a} & \text{при } -\frac{a}{2} \leqslant x \leqslant \frac{a}{2}; \\ 0 & \text{при } x > \frac{a}{2}. \end{cases}$$

Математическое ожидание и дисперсия погрешности округления в соответствии с (3.7) и (3.8)

$$a(x) = \int_{-\infty}^{\infty} x w_1(x)\,dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} x\frac{1}{a}\,dx = \frac{1}{a}\frac{x^2}{2}\Bigg|_{-\frac{a}{2}}^{\frac{a}{2}} = \frac{1}{a}\left(\frac{a^2}{8} - \frac{a^2}{8}\right) = 0;$$

$$D(x) = \int_{-\infty}^{\infty} [x - a(x)]^2 w_1(x)\,dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} x^2\frac{1}{a}\,dx = \frac{1}{a}\frac{x^3}{3}\Bigg|_{-\frac{a}{2}}^{\frac{a}{2}} = \frac{1}{a}\left(\frac{a^3}{24} + \frac{a^3}{24}\right) = \frac{a^2}{12}.$$

Пример 3.6. Случайная величина x распределена по нормальному закону, выражаемому функцией

$$w_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Показать, что a есть математическое ожидание, а σ^2 — дисперсия случайной величины x.

В соответствии с (3.7) математическое ожидание $a(x)$ случайной величины x

$$a(x) = \int_{-\infty}^{\infty} x w_1(x)\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx.$$

Введем новую переменную

$$z = \frac{x-a}{\sigma}.$$

Отсюда

$$x = \sigma z + a; \quad dx = \sigma\, dz.$$

Тогда получим

$$a(x) = \frac{\sigma}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sigma z + a) \exp\left\{-\frac{z^2}{2}\right\} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma z \exp\left\{-\frac{z^2}{2}\right\} dz + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz.$$

Первое слагаемое правой части равно нулю, так как под знаком интеграла нечетная функция. Известно, что интеграл

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \sqrt{2\pi},$$

тогда

$$a(x) = a.$$

Дисперсия в соответствии с (3.8)

$$D(x) = \int_{-\infty}^{\infty} [x - a(x)]^2 w_1(x)\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-a)^2 \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx.$$

Введем новую переменную

$$z = \frac{x-a}{\sigma}.$$

Отсюда

$$x - a = \sigma z; \quad dx = \sigma\, dz.$$

Тогда

$$D(x) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp\left\{-\frac{z^2}{2}\right\} dz.$$

Интегрируя по частям, положив $u = z$ и $z \exp\left\{-\frac{z^2}{2}\right\} dz = dv$, после определенных преобразований, получим

$$D(x) = \sigma^2.$$

Пример 3.7. Контрольно-измерительное устройство имеет систематическую погрешность +30 мВ и случайную погрешность, распределенную по нормальному закону со среднеквадратическим отклонением 10 мВ. Найти вероятность того, что общая погрешность устройства примет значение, принадлежащее интервалу 10 — 50 мВ.

Решение. Вероятность того, что погрешность x устройства не превысит по абсолютной величине заданного предела, равна

$$p(10 \leqslant x \leqslant 50) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{10}^{50} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx.$$

Введем новую переменную

$$z = \frac{x-a}{\sigma}.$$

Отсюда

$$x = \sigma z + a; \quad dx = \sigma dz.$$

Тогда

$$p(10 \leqslant x \leqslant 50) = \frac{1}{\sqrt{2\pi}} \int_{\frac{10-a}{\sigma}}^{\frac{50-a}{\sigma}} \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{1}{\sqrt{2\pi}} \int_{\frac{10-a}{\sigma}}^{0} \exp\left\{-\frac{z^2}{2}\right\} dz +$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{50-a}{\sigma}} \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{50-a}{\sigma}} \exp\left\{-\frac{z^2}{2}\right\} dz -$$

$$- \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{10-a}{\sigma}} \exp\left\{-\frac{z^2}{2}\right\} dz.$$

Пользуясь функцией Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} \exp\left\{-\frac{z^2}{2}\right\} dz,$$

окончательно получим

$$p(10 \leqslant x \leqslant 50) = \Phi\left(\frac{50-a}{\sigma}\right) - \Phi\left(\frac{10-a}{\sigma}\right).$$

В нашем случае

$$a = 30 \text{ мВ и } \sigma = 10 \text{ мВ}.$$

Тогда с помощью таблицы функции Лапласа (см. приложение) находим

$$p(10 \leqslant x \leqslant 50) = \Phi\left(\frac{50-30}{10}\right) - \Phi\left(\frac{10-30}{10}\right) =$$

$$= 2\Phi(2) = 0{,}4472 = 0{,}9544.$$

Пример 3.8. Цель может быть с равной вероятностью поражена в любой точке радиуса a. Определить математическое ожидание и дисперсию отклонения точки поражения от центра окружности.

Решение. Случайное расстояние r точки поражения от центра круга (рис. 3.31) может быть выражено через прямоугольные координаты x и y

$$r = \sqrt{x^2 + y^2}.$$

Исходя из того что вероятность попадания в окружность радиуса a равна единице, получим:

$$p[|x| \leqslant a; \quad |y| \leqslant a] = \iint\limits_S w_2(x, y)\, dx\, dy = w_2(x, y) \iint\limits_S dx\, dy = 1.$$

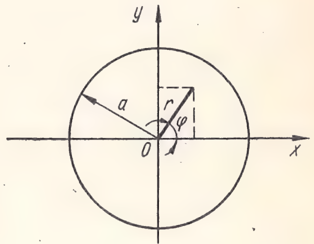


Рис. 3.31

Так как двойной интеграл $\iint\limits_S dx\, dy$ выражает площадь окружности радиуса a и равен πa^2, то совместная плотность вероятности случайных величин x и y аналитически может быть выражена следующим образом:

$$w_2(x, y) = \begin{cases} \dfrac{1}{\pi a^2} \text{ при } x^2 + y^2 \leqslant a^2; \\ 0 \text{ при } x^2 + y^2 > a^2. \end{cases}$$

Тогда математическое ожидание и дисперсия соответственно

$$a(r) = \iint\limits_{x^2+y^2 \leqslant a^2} \frac{1}{\pi a^2} \sqrt{x^2 + y^2}\, dx\, dy = \frac{1}{\pi a^2} \int\limits_0^{2\pi} d\varphi \int\limits_0^a r^2\, dr = \frac{2}{3} a;$$

$$D(r) = \iint\limits_{x^2+y^2 \leqslant a^2} \frac{1}{\pi a^2} (x^2 + y^2)\, dx\, dy - a^2(r) = \frac{1}{\pi a^2} \int\limits_0^{2\pi} d\varphi \int\limits_0^a r^3\, dr - \frac{4}{9} a^2 = \frac{a^2}{18}.$$

Пример 3.9. Корреляционная функция случайного процесса $x(t)$ задана выражением

$$K_x(\tau) = A \exp\{-\alpha(\tau)\},$$

где $\alpha > 0$ (рис. 3.32). Найти спектральную плотность.

Решение. В соответствии с (3.41) спектральная плотность

$$G(\omega) = \frac{2}{\pi} \int\limits_0^\infty K_x(\tau) \cos \omega\tau\, d\tau = \frac{2A}{\pi} \int\limits_0^\infty \exp\{-\alpha\tau\} \cos \omega\tau\, d\tau = \frac{2A\alpha}{\pi(\alpha^2 + \omega^2)}.$$

График $G(\omega)$ показан на рис. 3.33.

Пример 3.10. Найти спектральную плотность для процесса с корреляционной функцией вида δ-функции.

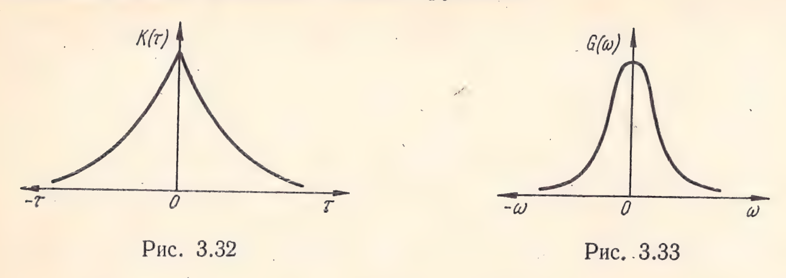


Рис. 3.32 Рис. 3.33

Решение. Исходя из определения δ-функции как предела прямоугольного импульса $\delta_\tau(t)$, длительности τ и высоты $\frac{1}{\tau}$ при $\tau \to 0$ интеграл $\int\limits_{-\infty}^{\infty} x(t)\,\delta(t)\,dt$, где $x(t)$ — произвольная функция, может быть представлен в виде

$$\int\limits_{-\infty}^{\infty} x(t)\,\delta(t)\,dt = \lim_{\tau\to 0}\int\limits_{-\infty}^{\infty} x(t)\,\delta_\tau(t)\,dt = \lim_{\tau\to 0}\int\limits_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)\,\frac{1}{\tau}\,dt.$$

Согласно теореме о среднем значении,

$$\lim_{\tau\to 0}\int\limits_{-\frac{\tau}{2}}^{\frac{\tau}{2}} x(t)\,\frac{1}{\tau}\,dt = \lim_{\substack{\tau\to 0\\ |\theta|\leqslant\tau}} x(0)\,\frac{1}{\tau}\int\limits_{-\frac{\tau}{2}}^{\frac{\tau}{2}} dt = x(0).$$

Таким образом,

$$\int\limits_{-\infty}^{\infty} x(t)\,\delta(t)\,dt = x(0). \qquad (3.75)$$

Тогда, согласно общей формуле (3.41), пользуясь (3.54), найдем спектральную плотность

$$G(\omega) = \frac{1}{\pi}\int\limits_{-\infty}^{\infty} K_x(\tau)\exp\{-j\omega\tau\}\,d\tau = \frac{1}{\pi}\int\limits_{-\infty}^{\infty} \delta(t)\exp\{-j\omega\tau\}\,dt =$$

$$= \frac{1}{\pi}\,|\exp\{-j\omega\tau\}\,|_{\tau=0} = \frac{1}{\pi}.$$

Следовательно, при корреляционной функции типа δ-функции спектр равномерен на всех частотах (сигнал типа «белый шум», см. рис. 3.8, *а* и *б*).

Пример 3.11. Найти корреляционную функцию для процесса с спектральной плотностью в виде δ-функции — $G(\omega) = \delta(\omega)$ (рис. 3.34).

Решение. Согласно общей формуле (3.42), пользуясь (3.54), получим

$$K(\tau) = \frac{1}{2}\int_{-\infty}^{\infty} G(\omega)\exp\{j\omega\tau\}\,d\omega = \frac{1}{2}\int_{-\infty}^{\infty}\delta(\omega)\exp\{j\omega\tau\}\,d\omega =$$

$$= \frac{1}{2}\,|\exp\{j\omega\tau\}|_{\tau=0} = \frac{1}{2}.$$

График корреляционной функции представлен на рис. 3.35.

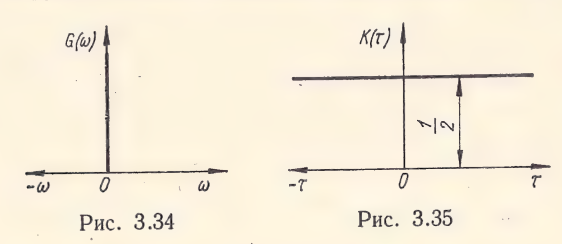


Рис. 3.34 Рис. 3.35

Пример 3.12. Найти корреляционную функцию периодического сигнала $x(t)$ с периодом $T_1 = \frac{2\pi}{\omega}$.

Решение. Корреляционная функция сигнала $x(t)$ равна

$$K_x(\tau) = m_1[x(t)\cdot x(t+\tau)] = \frac{1}{T_1}\int_{-\frac{T_1}{2}}^{\frac{T_1}{2}} x(t)\,x(t+\tau)\,dt.$$

Представим периодическую функцию разложением в ряд Фурье

$$x(t) = \sum_{k=-\infty}^{\infty}\dot{A}_k\exp\{jk\omega_1 t\}.$$

Тогда

$$K_x(\tau) = \frac{1}{T_1}\int_{-\frac{T_1}{2}}^{\frac{T_1}{2}}\sum_{k=-\infty}^{\infty}\dot{A}_k\exp\{jk\omega_1 t\}\cdot\sum_{l=-\infty}^{\infty}\dot{A}_l\exp\{jl\omega_1(t+\tau)\}\,dt =$$

$$= \frac{1}{T_1}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\dot{A}_k\dot{A}_l\exp\{jl\omega_1\tau\}\int_{-\frac{T_1}{2}}^{\frac{T_1}{2}}\exp\{j(k+l)\,\omega_1 t\}\,dt =$$

$$= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}A_kA_l\,\frac{\sin(k+l)\,\omega_1\frac{T_1}{2}}{j(k+l)\,\omega_1\frac{T_1}{2}}\exp\{jl\omega_1\tau\}.$$

Так как $\omega_1 \frac{T_1}{2} = \pi$, то дробь $\dfrac{\sin (k+l)\,\omega_1 \frac{T_1}{2}}{j\,(k+l)\,\omega_1 \frac{T_1}{2}}$ равна нулю при любых комбинациях k и l, кроме $k+l=0$, при которой дробь обращается в единицу. Следовательно,

$$K_x(\tau) = \sum_{k=-\infty}^{\infty} \dot{A}_к \dot{A}_{-к} \exp\{jk\omega_1\tau\} = \sum_{k=-\infty}^{\infty} |\dot{A}_к|^2 \exp\{jk\omega_1\tau\} =$$

$$= 2\sum_{k=1}^{\infty} |\dot{A}_к|^2 \cos k\omega_1\tau.$$

Как видно из полученного выражения, корреляционная функция периодического сигнала с периодом T_1 представляет собой периодическую функцию аргумента τ с тем же периодом T_1. Амплитуда гармоники корреляционной функции равна удвоенному квадрату амплитуды соответствующей гармоники $x(t)$. При этом $K_x(\tau)$ как четная функция разлагается в ряд по косинусам.

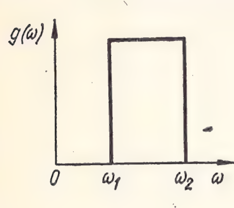


Рис. 3.36

Пример 3.13. Нормированная спектральная плотность $g(\omega)$ случайной функции $x(t)$ постоянна на некотором интервале частот ω_1, ω_2 и равна нулю вне этого интервала (рис. 3.36). Определить нормированную корреляционную функцию.

Решение. Исходя из условия, что площадь, ограниченная кривой $g(\omega)$, равна единице, находим значение $g(\omega)$ при $\omega_1 \leqslant \omega \leqslant \omega_2$

$$g(\omega)(\omega_2 - \omega_1) = 1;$$

$$g(\omega) = \frac{1}{\omega_2 - \omega_1}.$$

Из (3.42) имеем

$$r(\tau) = \int_{\omega_1}^{\omega_2} g(\omega)\cos\omega\tau\, d\omega = \frac{1}{\omega_2 - \omega_1}\int_{\omega_1}^{\omega_2} \cos\omega\tau\, d\omega =$$

$$= \frac{1}{\tau(\omega_2 - \omega_1)}(\sin\omega_2\tau - \sin\omega_1\tau) =$$

$$= \frac{2}{\tau(\omega_2 - \omega_1)}\cos\left(\frac{\omega_2 + \omega_1}{2}\tau\right)\sin\left(\frac{\omega_2 - \omega_1}{2}\tau\right).$$

Общий вид функции $r(\tau)$ изображен на рис. 3.37.

Пример 3.14. Белый шум со спектральной плотностью $G(\omega) = G_0 = \text{const}$ проходит через многокаскадный усилитель с частотной характеристикой в форме гауссовой кривой (рис. 3.38)

$$G(\omega) = G_0 \exp\left\{-\frac{(\omega - \omega_0)^2}{2\beta^2}\right\}.$$

Найти спектральную плотность и корреляционную функцию выходного сигнала.

Решение. Спектральная плотность выходного сигнала

$$G_1(\omega) = G_0 \,|\, C(\omega)\,|^2 = G_0 C_0^2 \exp\left\{-\frac{(\omega-\omega_0)^2}{\beta^2}\right\}.$$

Таким образом, спектр шумов на выходе усилителя также имеет форму гауссовой кривой.

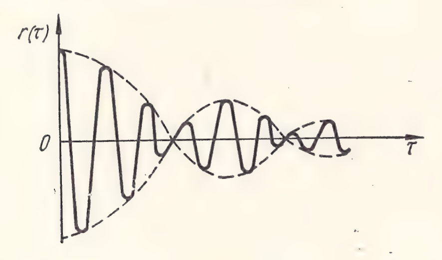


Рис. 3.37

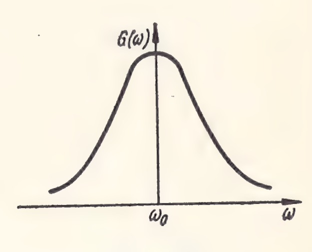


Рис. 3.38

Для случая, когда спектр сигнала узкополосный, симметричен относительно частоты ω_0 и обладает в этой точке максимумом, корреляционная функция в соответствии с (3.50)

$$K(\tau) = [2G_1(\omega_0 - \omega)\cos\omega\tau\, d\omega]\cos\omega_0 t =$$

$$= 2G_0C_0^2\cos\omega_0\tau\int_0^\infty \exp\left\{-\frac{\omega^2}{\beta^2}\right\}\cos\omega\tau\, d\omega =$$

$$= \sqrt{\pi}\, G_0C_0^2\beta\exp\left\{-\frac{\beta^2\tau^2}{4}\right\}\cos\omega_0\tau.$$

Пример 3.15. Белый шум U_0 со спектральной плотностью $G(\omega) = G_0 = \mathrm{const}$ подается на вход цепи, изображенной на рис. 3.39. Найти спектральную плотность, корреляционную функцию и среднеквадратическое значение выходного сигнала U_1.

Рис. 3.39

Решение. Спектральная плотность выходного сигнала

$$G(\omega) = G_0\,|\,K(j\omega)\,|^2,$$

где $K(j\omega)$ — коэффициент передачи цепи.

Для схемы, изображенной на рис. 3.39,

$$K(j\omega) = \frac{R_1}{R_1 + R_2 + j\omega L} = \frac{\alpha}{1 + j\omega T},$$

где

$$\alpha = \frac{R_1}{R_1 + R_2};\quad T = \frac{L}{R_1 + R_2}.$$

Квадрат модуля коэффициента передачи

$$|K(j\omega)|^2 = \frac{\alpha^2}{1+(\omega T)^2}.$$

Тогда спектральная плотность выходного сигнала

$$G(\omega) = G_0 |K(j\omega)|^2 = \frac{G_0\alpha^2}{1+(\omega T)^2}.$$

Корреляционная функция в соответствии с (3.42)

$$K(\tau) = \int_0^\infty G_1(\omega)\cos\omega\tau\, d\omega = G_0\alpha^2 \int_0^\infty \frac{\cos\omega\tau}{1+(\omega T)^2}\, d\omega.$$

Введя новую переменную $x = \omega T$, получим

$$K(\tau) = \frac{G_0\alpha^2}{T}\int_0^\infty \frac{\cos\frac{\tau}{T}x}{1+x^2}\, dx.$$

Известно, что

$$\int_0^\infty \frac{\cos px}{q^2+x^2}\, dx = \frac{\pi}{2q}\exp\{-|pq|\}.$$

Тогда

$$K(\tau) = \frac{G_0\,\alpha^2}{T}\,\frac{\pi}{2}\exp\left\{-\left|\frac{\tau}{T}\right|\right\} = \frac{\pi G_0\alpha^2}{2T}\exp\left\{-\frac{|\tau|}{T}\right\}.$$

Среднеквадратическое значение выходного сигнала

$$\sigma = \sqrt{K(\tau)_{\tau=0}} = \sqrt{\frac{\pi G_0\alpha^2}{2T}}.$$

Пример 3.16. На вход дифференцирующего устройства поступает случайная функция $x(t)$ с математическим ожиданием $a_x(t) = \sin t$ и корреляционной функцией $K_x(t_2, t_1) = D(x)\exp\{-\alpha(t_2 - t_1)^2\}$, где $D(x)$ — постоянная во времени дисперсия функции $x(t)$. Определить математическое ожидание и дисперсию на выходе устройства.

Решение. Случайная функция $y(t)$ на выходе устройства связана с входной функцией $x(t)$ оператором дифференцирования

$$y(t) = \frac{d}{dt}x(t).$$

Тогда математическое ожидание и дисперсия выходного сигнала

$$a_y(t) = \frac{d}{dt}a_x(t) = \cos t;$$

$$K_y(t_2, t_1) = \frac{\partial^2}{\partial t_1\,\partial t_2}K_x(t_2, t_1) = 2D(x)\alpha\exp\{-\alpha(t_2 - t_1)^2\}[1 + 2\alpha(t_2 - t_1)^2].$$

При

$$t_2 = t_1$$

$$K_y(t_2 - t_1) = D_y(t) = 2D(x)\alpha.$$

Таким образом, дисперсия выходного сигнала также не зависит от времени.

Пример 3.17. Определить погрешность квантования во времени сигнала конечной длительности. Известно следующее:

1) сигнал $x(t)$ может принимать с равной вероятностью любые значения в пределах от 0 до $x_{\text{макс}}$;

2) частота квантования $\omega_к$ выбрана таким образом, что относительная величина площади энергетического спектра сигнала, находящегося в пределах частот от $\omega_c = \frac{1}{2}\omega_к$ до $\omega = \infty$, равна 5%.

Решение. Для решения задачи воспользуемся выражением (3.66). По условию относительная величина площади отсекаемого участка спектра сигнала $Q_{\omega_c} = 0{,}05$. Средняя мощность сигнала $P_T = \int\limits_0^{x_{\text{макс}}} x^2 \omega(x)\,dx.$

Так как сигнал по шкале уровней распределен равновероятно, то $\omega(x) = \frac{1}{x_{\text{макс}}}$. Тогда

$$P_T = \int\limits_0^{x_{\text{макс}}} x^2 \frac{1}{x_{\text{макс}}}\,dx = \frac{x^2_{\text{макс}}}{3}.$$

Следовательно, искомая погрешность квантования

$$D_{\omega_c} = \frac{0{,}05x^2_{\text{макс}}}{3 \cdot x^2_{\text{макс}}} \cdot 100\% = 1{,}7\%.$$

Пример 3.18. Передаваемый по каналу связи сигнал квантуется по уровню способом замены его мгновенных значений ближайшим меньшим квантованным уровнем. Определить необходимое количество уровней квантования сигнала при условии, что приведенная среднеквадратическая погрешность квантования не превышает 0,3%.

Решение. При заданном способе квантования погрешность квантования отрицательная и может принимать значения от 0 до $-\Delta_x$, где Δ_x — шаг квантования.

Среднеквадратическое значение погрешности квантования в соответствии с (3.72)

$$\sigma_к = \frac{\Delta_x}{2\sqrt{3}}.$$

Приведенная среднеквадратическая погрешность

$$\gamma_к\% = \frac{\sigma_к}{x_{\text{макс}}} \cdot 100 = \frac{\Delta_x}{2\sqrt{3}N\Delta_x} \cdot 100 = \frac{1}{2\sqrt{3}N} \cdot 100,$$

равен $\Delta_x = t_0 = \frac{1}{f_0}$. Так как нет синхронизации между посылками импульсов эталонного генератора и началом временного интервала t_x, то имеем по существу случай квантования, когда дискретные уровни не фиксированы относительно нулевого уровня входного сигнала (рис. 3.27). При этом погрешность квантования распределена по закону Симпсона, а среднеквадратическое значение этой погрешности

$$\sigma_k = \frac{\Delta_x}{\sqrt{6}} = \frac{1}{\sqrt{6}},$$

так как шаг квантования Δ_x, приведенный к выходу устройства, равен единице (одному импульсу). Тогда общая среднеквадратическая погрешность преобразователя

$$\sigma_\Sigma = \sqrt{(f_0\sigma_{ВИП})^2 + (t_x\sigma_{f_0})^2 + \frac{1}{6}}.$$

Переходя к относительным величинам, получим

$$\gamma_\Sigma\% = \frac{\sigma_\Sigma}{n}100 = \sqrt{\left(\frac{\sigma_{ВИП}}{t_x}\right)^2 + \left(\frac{\sigma_{f_0}}{f_0}\right)^2 + \frac{1}{6}\left(\frac{100}{K_{ВИП}f_0U_x}\right)^2} =$$

$$= \sqrt{(0,1)^2 + (0,05)^2 + \frac{1}{6}\left(\frac{100}{10^{-4}10^5 10^2}\right)^2} = 0,12\%.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие трудности возникают при использовании n-мерных плотностей распределения вероятностей случайного процесса для анализа систем передачи информации?
2. Какими усредненными характеристиками описываются обычно случайные процессы?
3. Что понимается под математическим ожиданием и дисперсией случайного процесса?
4. Что понимается под среднеквадратическим значением и среднеквадратическим отклонением случайной величины?
5. Что понимается под корреляционной функцией случайного процесса?
6. Что понимается под стационарностью случайного процесса в широком и узком смыслах; всегда ли процессы, стационарные в узком смысле, будут стационарными в широком смысле, и наоборот?
7. Каковы основные свойства корреляционной функции стационарного процесса?
8. Что понимается под временем корреляции случайного процесса?
9. Какие случайные процессы называются эргодическими?
10. Почему невозможно непосредственное приложение классического гармонического анализа к случайным процессам?
11. Какова связь между спектральной плотностью и корреляционной функцией случайного процесса?
12. Какой случайный процесс называют белым шумом?
13. Какова корреляционная функция белого шума?
14. Что понимается под эффективной шириной спектра случайного процесса?
15. Какова связь между интервалом корреляции и эффективной шириной спектра случайного процесса?
16. В чем сущность квантования непрерывного сигнала по времени?
17. Что такое функция отсчетов и каков ее вид?

18. Сформулируйте критерий Котельникова, в чем значение этого критерия для теории и техники передачи сообщений?

19. В чем заключаются противоречия критерия Котельникова?

20. Сформулируйте критерий Железнова, сопоставьте критерий Железнова с критерием Котельникова.

21. В чем сущность квантования сигналов по уровню и в каких случаях прибегают к такому квантованию?

22. Какие известны способы квантования сигналов по уровню?

23. В каких случаях прибегают к неравномерному квантованию по уровню?

Глава IV

ИНФОРМАЦИОННЫЕ МОДЕЛИ СИГНАЛОВ

Всякое информационное передающее устройство работает по единой схеме (рис. 4.1). На вход устройства подается совокупность сообщений $x_1, x_2, \ldots, x_n$. Задача устройства состоит в том, чтобы передать совокупность этих сообщений с достаточно высокой достоверностью, или, иными словами, чтобы перевести вектор сообщений на входе $X(x_1, x_2, \ldots, x_n)$ в соответствующий ему вектор сообщений на выходе $Y(y_1, y_2, \ldots, y_n)$ без ошибок или с допустимыми ошибками [23, 24, 25, 26, 27].

Рис. 4.1

В процессе передачи сообщение может подвергаться многочисленным преобразованиям, существенно меняющим его физические характеристики. Однако передаваемая информация должна оставаться инвариантной при всех преобразованиях. Естественно, что количество передаваемой получателю информации связано с неопределенностью, которая имела место относительно передаваемого сообщения. В связи с этим необходимо ввести количественную меру оценки информации и неопределенности передаваемых сообщений.

4.1. Количество информации и неопределенность. Энтропия как мера неопределенности

Фундаментальным вопросом для теории информации является вопрос о количественной мере информации. Необходимо отметить, что всякая информация получается потребителем после принятия сообщения, т. е. в результате опыта. Сообщение, получаемое на приемной стороне, несет полезную информацию лишь в том случае, если имеется неопределенность относительно состояния источника сообщений.

Если опыт имеет лишь один исход и не содержит никакой неопределенности, то наблюдатель заранее будет знать исход этого опыта. В результате осуществления такого опыта наблюдатель не получит никакой информации.

Пусть опыт имеет два равновероятных исхода. Результат контроля, например, должен указать, что контролируемый параметр находится в пределах нормы или за ее пределами. Передаваемое сообщение в этом случае может принимать два значения и содержит определенную информацию.

Рассмотрим в качестве третьего примера источник напряжения, выходное напряжение которого может с одинаковой вероятностью принимать десять различных значений. В этом случае будет большая предварительная неопределенность относительно источника, а поступившее сообщение (конкретный исход опыта) даст более уточненную характеристику состояния входа источника.

Рассмотрим общий случай, когда источник может передавать n независимых и несовместных сообщений $x_1, x_2, \ldots, x_n$ с вероятностями $p(x_1), p(x_2), \ldots, p(x_n)$ соответственно. Естественно, что чем меньшая априорная вероятность события, тем большее количество информации оно несет. Так, например, сообщение о том, что летом температура воздуха в Крыму выше нуля, не несет существенной информации, ибо вероятность такого события очень велика. Сообщение же о том, что температура воздуха на южном берегу Крыма в июне ниже нуля, содержит значительно большее количество информации, ибо такое событие является весьма редким.

Поэтому естественно предположить, что количественной мерой неопределенности отдельного сообщения, а также передаваемой им информации может быть величина, обратная его априорной вероятности $\frac{1}{p(x_i)}$. Однако такая мера неудобна тем, что в случае, когда опыт будет иметь только один исход, т. е. вероятность такого события равна единице, количество информации, согласно принятой мере, равно единице. В действительности же результат такого опыта не дает никакой информации. Кроме того, такая мера не обладает свойством аддитивности. Действительно, если имеет место сложное событие, состоящее из двух независимых событий x_i и x_j, то вероятность такого события будет определяться произведением вероятностей $p(x_i) \cdot p(x_j)$. Количество информации в сложном сообщении должно оцениваться величиной $\frac{1}{p(x_i)\, p(x_j)}$. С этой точки зрения более удобной является логарифмическая мера количества информации

$$I(x_i) = \log_a \frac{1}{p(x_i)}. \tag{4.1}$$

При этом количество информации, содержащееся в сложном сообщении, представляющем совокупность событий x_i и x_j, будет равно

$$I(x_i,\ x_j) = \log_a \frac{1}{p(x_i)\, p(x_j)} = \log_a \frac{1}{p(x_i)} + \\ + \log_a \frac{1}{p(x_j)} = I(x_i) + I(x_j).$$

Логарифмическая мера, как видно, обладает свойством аддитивности. Кроме того, эта мера в случае событий с одним исходом дает нулевое количество информации.

Выражение (4.1) характеризует количество информации, содержащееся в сообщении x_i. Оно характеризует также априорную неопределенность этого сообщения. В связи с этим выражение (4.1) может быть использовано для количественной оценки неопределенности сообщения

$$H(x_i) = \log_a \frac{1}{p(x_i)}. \qquad (4.2)$$

Величину (4.2), характеризующую неопределенность отдельного (i-го) сообщения, принято называть частной энтропией.

Количество информации и неопределенность для всей совокупности случайных сообщений можно получить усреднением по всем событиям

$$I(X) = \sum_{i=1}^{n} p(x_i) \log_a \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_a p(x_i), \qquad (4.3)$$

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_a p(x_i). \qquad (4.4)$$

Зависимости (4.3) и (4.4) выражают среднее на одно событие (сообщение) количество информации и энтропии. Термин «энтропия» заимствован из термодинамики, где аналогичное выражение характеризует среднюю неопределенность состояния системы молекул вещества.

Несмотря на совпадение зависимостей (4.3) и (4.4), энтропия $H(X)$ и количество информации $I(X)$ принципиально различны. $H(X)$, выражающая среднюю неопределенность состояния источника сообщений, является объективной характеристикой источника сообщений и, если известна статистика сообщений, может быть вычислена априорно, т. е. до получения сообщения. Величина же $I(X)$ определяется апостериорно, т. е. после получения сообщений. $H(X)$ есть мера недостатка информации о состоянии отдельной системы. С поступлением информации о состоянии системы энтропия последней снижается.

Совпадение выражения (4.3) с выражением (4.4) свидетельствует лишь о том, что количество получаемой информации равно численно энтропии, которая имела место относительно источника сообщений.

На рассмотренной взаимосвязи количества информации с энтропией проявляется известный диалектический закон единства и борьбы противоположностей, так как информация рассматривается в связи со своей противоположностью — энтропией и, с другой стороны, рассматривается как мера уничтожения, снятия энтропии.

Единицы измерения количества информации и энтропии зависят от выбора основания логарифма в формулах (4.3) и (4.4). При использовании десятичных логарифмов количество информации и энтропия определяются в десятичных единицах — дитах. В случае

использования двоичных логарифмов измеряются количество информации и энтропия в двоичных единицах — битах. Наконец, при использовании натуральных логарифмов единицей измерения является натуральная единица — нит.

При анализе информационных процессов в электронных вычислительных машинах и других устройствах, функционирующих на основе двоичной системы счисления, удобно пользоваться двоичными единицами. При анализе процессов в приборах, работающих в десятичной (или двоично-десятичной) системе счисления, целесообразно пользоваться дитами. В математических выкладках удобно пользоваться натуральными единицами. При дальнейшем изложении будем пользоваться двоичными единицами — дв. ед.

Мера количества информации в виде (4.3) впервые была предложена К. Шенноном в 1948 году и затем более строго определена А. Я. Хинчиным [31].

В случае равной вероятности сообщений выражение (4.3) для количества информации приводится к виду

$$I(X) = -\log_2 p(x_i) = \log_2 n, \qquad (4.5)$$

где $n = \frac{1}{p(x_i)}$ — количество передаваемых сообщений.

Такая мера количества информации была предложена в 1928 году Р. Хартли.

4.2. Свойства энтропии дискретных сообщений

Формула (4.4) выражает энтропию дискретных сообщений.

Энтропия дискретных сообщений обладает следующими свойствами:

1. Энтропия есть величина вещественная, ограниченная и неотрицательная. Это свойство следует из выражения (4.4), если учесть, что вероятности $p(x_i)$ есть величины неотрицательные, заключенные в промежутке $0 \leqslant p(x_i) \leqslant 1$.

2. Энтропия детерминированных сообщений равна нулю. Действительно, если заранее известно, какое будет событие, например x_1, то вероятность этого события равна единице, а остальных — нулю, т. е. $p(x_1) = 1$; $p(x_2) = p(x_3) = \cdots = p(x_n) = 0$.

Запишем выражение (4.4) в виде

$$H(X) = -\left[p(x_1)\log_2 p(x_1) + \sum_{i=2}^{n} p(x_i)\log_2 p(x_i)\right].$$

Первый член равен нулю, поскольку $\log_2 1 = 0$. Остальные члены также обращаются в нуль при $p(x_i) \to 0$. В самом деле, устремляя $p(x_i)$ к нулю, получим

$$\lim_{p(x_i)\to 0} [-p(x_i)\log_2 p(x_i)] = \lim_{p(x_i)\to 0} \frac{\log_2 \frac{1}{p(x_i)}}{\frac{1}{p(x_i)}}.$$

Введя замену $\frac{1}{p(x_i)} = \beta$ и раскрывая неопределенность по правилу Лопиталя, получим

$$\lim_{\beta \to \infty} \frac{\log_2 \beta}{\beta} = \lim_{\beta \to \infty} \frac{\frac{1}{\beta} \log_2 \beta}{1} = 0.$$

3. Энтропия максимальна, если все события равновероятны.

Нахождение оптимального значения вероятностей $p(x_i)$, при котором обеспечивается максимум энтропии, может быть произведено методом Лагранжа для отыскания условного экстремума. Для этого необходимо составить вспомогательную функцию F вида

$$F = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) - \lambda,$$

где λ — множитель Лагранжа.

С учетом того, что $\sum_{i=1}^{n} p(x_i) = 1$, представим вспомогательную функцию в виде

$$F = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) - \lambda \sum_{i=1}^{n} p(x_i) =$$

$$= -\sum_{i=1}^{n} [p(x_i) \log_2 p(x_i) + \lambda p(x_i)] = \sum_{i=1}^{n} F_i$$

и найдем максимум этой функции.

Искомый максимум будет иметь место, когда

$$\frac{\partial F_i}{\partial p(x_i)} = -\log_2 p(x_i) - \log_2 e - \lambda = 0 \quad (i = 1, 2, \ldots, n)$$

или

$$\log_2 p(x_i) = -\log_2 e - \lambda \quad (i = 1, 2, \ldots, n).$$

Величина вероятности $p(x_i)$ не зависит от номера i, что может быть только тогда, когда все $p(x_i)$ равны между собой.

Следовательно, энтропия максимальна при

$$p(x_1) = p(x_2) = \cdots = p(x_n) = \frac{1}{n},$$

что и требовалось доказать.

Максимальное значение энтропии

$$H(X)_{\text{макс}} = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n. \qquad (4.6)$$

Как видно из выражения (4.6), в случае равновероятных событий энтропия возрастает с увеличением количества событий.

4. Энтропия системы двухальтернативных событий может изменяться в пределах от нуля до единицы.

Энтропия системы двух событий

$$H(X) = -p(x_1)\log_2 p(x_1) - p(x_2)\log_2 p(x_2) =$$
$$= -p(x_1)\log_2 p(x_1) - [1 - p(x_1)]\log_2 [1 - p(x_1)].$$

Из последнего выражения видно, что энтропия равна нулю при

$$p(x_1) = 0;\ p(x_2) = 1,$$

или

$$p(x_1) = 1;\ p(x_2) = 0.$$

Максимум энтропии будет иметь место, когда

$$p(x_1) = p(x_2).$$

При этом максимальное значение энтропии

$$H(X)_{\text{макс}} = -\frac{1}{2}\log_2\frac{1}{2} - $$
$$-\frac{1}{2}\log_2\frac{1}{2} = 1 \text{ дв. ед.}$$

Таким образом, можно утверждать, что одна двоичная единица — это энтропия системы двух равновероятных независимых событий.

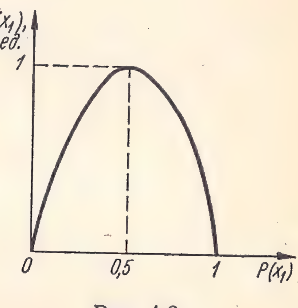


Рис. 4.2

На рис. 4.2 представлена графически зависимость энтропии бинарных событий от вероятности одного из событий.

4.3. Энтропия непрерывных сообщений

Непрерывное сообщение, как случайная величина, характеризуется дифференциальным законом распределения вероятностей $w(x)$.

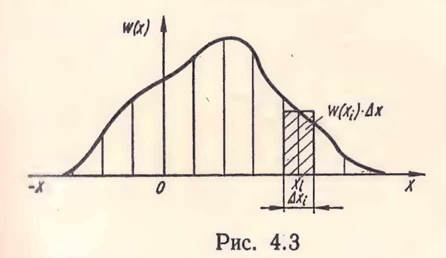


Рис. 4.3

Пусть функция распределения плотности вероятности непрерывного сообщения имеет вид (рис. 4.3).

Для определения энтропии непрерывных сообщений воспользуемся в качестве исходного выражения (4.4) для энтропии дискретных сообщений. В связи с этим разобьем шкалу уровней непрерывной случайной величины x на небольшие участки Δx_i и внутри каждого участка выберем точки x_i так, чтобы выполнялось условие

$$\int_{\Delta x_i} w(x)\,dx = w(x_i)\,\Delta x_i. \tag{4.7}$$

Выражение (4.7) характеризует вероятность $p(x_i)$ попадания случайной величины x в интервал Δx_i (рис. 4.3) при замене непрерывной случайной величины x совокупностью дискретных сообщений x_i, вероятности поступления которых определяются выражением (4.4). Такая замена будет тем точнее, чем меньше участки Δx_i.

Энтропия эквивалентного дискретного сообщения в соответствии с (4.4)

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) = -\sum_{i=1}^{n} w(x_i) \Delta x_i \log_2 [w(x_i) \Delta x_i] =$$
$$= -\sum_{i=1}^{n} w(x_i) \Delta x_i \log_2 w(x_i) - \sum_{i=1}^{n} w(x_i) \Delta x_i \log_2 \Delta x_i.$$

Чтобы получить выражение для энтропии непрерывного сообщения, осуществим предельный переход при $\Delta x_i \to 0$.

$$\lim_{\Delta x_i \to 0} \left\{-\sum_{i=1}^{n} w(x_i) \Delta x_i \log_2 w(x_i) - \sum_{i=1}^{n} w(x_i) \Delta x_i \log_2 \Delta x_i\right\} =$$
$$= -\int_{-\infty}^{\infty} w(x) \log_2 w(x)\, dx - \lim_{\Delta x_i \to 0} \log_2 \Delta x_i, \qquad (4.8)$$

так как

$$\lim_{\Delta x_i \to 0} \sum_{i=1}^{n} w(x_i) \Delta x_i = \int_{-\infty}^{\infty} w(x)\, dx = 1.$$

Второй член выражения (4.8) при $\Delta x_i \to 0$ стремится к бесконечности. Следовательно, энтропия непрерывного сообщения должна быть равна бесконечности. Однако в реальных условиях отсчет сообщений на приемной стороне производится в дискретных точках вследствие конечной точности и разрешающей способности аппаратуры, т. е. интервалы Δx_i имеют конечную величину. Поэтому выражение (4.8), определяющее энтропию непрерывного сообщения, имеет две составляющие, одна из которых определяется законом распределения сообщения, а вторая является постоянной величиной и обычно исключается из рассмотрения.

Первое слагаемое

$$h(x) = -\int_{-\infty}^{\infty} w(x) \log_2 w(x)\, dx \qquad (4.9)$$

представляет собой так называемую дифференциальную энтропию. Дифференциальная энтропия зависит от статистики сообщений. Если положить величину x безразмерной, то $h(x)$ можно измерять в двоичных единицах, приходящихся на один определяющий дискрет в смысле В. А. Котельникова [27]. Следует, однако, иметь в виду, что в отличие от энтропии дискретных сообщений величина $h(x)$ является относительной, так как зависит от масштаба x, а значит, от выбора

единицы измерения. В связи с этим $h(x)$ отдельно не может служить абсолютной мерой неопределенности непрерывного сообщения [9, 17].

Доказано [27], что при заданной средней мощности (дисперсии) максимальной энтропией обладает нормальный закон распределения вероятностей. Если же задана пиковая мощность, то максимальной энтропией обладает равновероятный закон распределения.

4.4. Энтропия сложных сообщений

При решении задач передачи информации часто имеют дело с несколькими источниками, дающими зависимые сообщения. Совокупность сообщений, вырабатываемых несколькими источниками, назовем сложным сообщением.

Пусть имеются два источника сообщений. Сообщения первого источника принимают значения $x_1, x_2, \ldots, x_n$ с вероятностями $p(x_1), p(x_2), \ldots, p(x_n)$ и сообщения второго источника принимают значения $y_1, y_2, \ldots, y_m$ с вероятностями $p(y_1), p(y_2), \ldots, p(y_m)$.

Совместную энтропию совокупности сообщений X и Y можно представить в виде

$$H(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 p(x_i, y_j), \tag{4.10}$$

где $p(x_i, y_j)$ — вероятность совместного появления сообщений x_i и y_j.

Учитывая, что совместная вероятность может быть представлена в виде $p(x_i, y_j) = p(x_i) \cdot p(y_j/x_i)$, где $p(y_j/x_i)$ — условная вероятность сообщения y_j при условии, что поступило сообщение x_i, выражение (4.10) можно привести к виду

$$H(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i) p(y_j/x_i) \log_2 [p(x_i) p(y_j/x_i)] =$$
$$= -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \sum_{j=1}^{m} p(y_j/x_i) -$$
$$- \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j/x_i) \log_2 p(y_j/x_i).$$

С учетом того, что $\sum_{j=1}^{m} p(y_i/x_i) = 1$, так как при наличии сообщения x_i обязательно будет одно из сообщений ансамбля y_j, выражение для совместной энтропии может быть преобразовано следующим образом:

$$H(X, Y) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) - \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j/x_i) \log_2 p(y_j/x_i) =$$
$$= H(X) + \sum_{i=1}^{n} p(x_i) H(Y/x_i), \tag{4.11}$$

где $H(Y/x_i) = -\sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j/x_i)$ — так называемая частная условная энтропия, выражающая энтропию сообщения Y при условии, что имело место сообщение X. Второй член выражения (4.11) представляет собой усреднение $H(Y/x_i)$ по всем сообщениям $x_1, x_2, \ldots, x_n$ и называется средней условной энтропией сообщения Y при условии поступления сообщения X. Обозначив его через $H(Y/X)$, окончательно получим

$$H(X, Y) = H(X) + H(Y/X), \qquad (4.12)$$

где

$$H(Y/X) = -\sum_{i=1}^{n} p(x_i) H(Y/x_i) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i) p(y_j/x_i) \times$$
$$\times \log_2 p(y_j/x_i) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(y_j/x_i).$$

Основной смысл условной энтропии $H(Y/X)$ состоит в том, что она показывает, какую энтропию дают сообщения Y, когда уже известна энтропия сообщений X.

Из очевидного равенства $H(X, Y) = H(Y, X)$ получим

$$H(X, Y) = H(Y) + H(X/Y),$$

где

$$H(X/Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i/y_j).$$

Таким образом, совместная энтропия двух сообщений равна сумме безусловной энтропии одного из сообщений и условной энтропии второго сообщения.

При некотором множестве сообщений $X, Y, Z, \ldots$ совместная энтропия

$$H(X, Y, Z, \ldots) = H(X) + H(Y, Z, \ldots/X) =$$
$$= H(Y) + H(X, Z, \ldots/Y) = H(Z) + H(X, Y, \ldots/Z) = \ldots$$

Можно отметить следующие основные свойства энтропии сложных сообщений:

1. При статистически независимых сообщениях X и Y совместная энтропия равна сумме энтропий сообщений

$$H(X, Y) = H(X) + H(Y). \qquad (4.13)$$

Действительно, при статистически независимых сообщениях условная вероятность равна безусловной $p(y_j/x_i) = p(y_j)$, а совместная вероятность равна произведению вероятностей $p(x_i, y_j) = p(x_i) \cdot p(y_j)$. Тогда условная энтропия

$$H(Y/X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i) p(y_j)\log_2 p(y_j).$$

Произведя суммирование по i, с учетом равенства $\sum_{i=1}^{n} p(x_i) = 1$, получим

$$H(Y/X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i)\, p(y_j) \log_2 p(y_j) = -\sum_{j=1}^{m} p(y_j) \log_a p(y_j) = H(Y).$$

Следовательно, при статистически независимых сообщениях условная энтропия событий равна безусловной.

2. При полной статистической зависимости сообщений X и Y совместная энтропия равна безусловной энтропии одного из сообщений.

Действительно, полная статистическая зависимость соответствует случаю, когда условные вероятности $p(y_j/x_i)$ и $p(x_i/y_j)$ равны нулю или единице. В этом случае выражения

$$p(y_j/x_i) \log_2 (y_j/x_i) = 0 \text{ и } p(x_i/y_j) \log_2 (x_i/y_j) = 0$$

и, следовательно, условные энтропии $H(Y/X) = 0$ и $H(X/Y) = 0$. Тогда получим

$$H(X, Y) = H(X) = H(Y). \tag{4.14}$$

3. Условная энтропия может изменяться в пределах

$$0 \leqslant H(Y/X) \leqslant H(Y).$$

Так как условная энтропия положительна, равна нулю при полной статистической зависимости событий, максимальна при полной статистической независимости событий и равна безусловной энтропии, то отсюда непосредственно вытекает это свойство.

4. Для совместной энтропии всегда справедливо соотношение

$$H(X, Y) \leqslant H(X) + H(Y). \tag{4.15}$$

Данное свойство совместной энтропии непосредственно вытекает из предыдущего соотношения.

Если X и Y являются непрерывными случайными величинами, то по аналогии с выражением для безусловной энтропии (4.9) выражение для энтропии объединения сообщений X и Y можно представить в виде

$$h(X, Y) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 (X, Y)\, dY =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(X)\, dX\, dY -$$

$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(Y/X)\, dX\, dY.$$

Произведя интегрирование в первом слагаемом по Y и учитывая, что

$$\int_{-\infty}^{\infty} w(X, Y)\, dY = w(X),$$

получим

$$h(X, Y) = -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX -$$

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X, Y) \log_2 w(Y/X)\, dX\, dY = h(X) + h(Y/X),$$

где $h(X) = -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX$ — дифференциальная энтропия сообщения X; $h(Y/X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X, Y) \log_2 w(Y/X)\, dX\, dY$ — условная дифференциальная энтропия сообщения Y; $w(X, Y)$ — плотность совместного распределения X и Y; $w(X)$ — плотность распределения X; $w(Y/X)$ — условная плотность распределения Y относительно X.

4.5. Количество информации при неполной достоверности сообщений

В реальных условиях передача сообщений происходит при воздействии помех. Помехи искажают сообщения, вследствие чего сообщения на приемной стороне будут в той или иной степени отличаться от переданных, т. е. будет иметь место неполная достоверность передачи. Оценим количество передаваемой при этом информации для случаев дискретных и непрерывных сообщений.

1. Дискретные сообщения

Вследствие отличия принимаемых сообщений от передаваемых, при оценке количества передаваемой информации целесообразнее рассматривать две системы: систему передаваемых X и систему принимаемых сообщений Y.

Пусть передаваемое сообщение может принимать значения x_1, x_2, ..., x_n с априорными вероятностями соответственно $p(x_1)$, $p(x_2)$, ..., $p(x_n)$.

Принимаемые сообщения характеризуются совокупностью значений y_1, y_2, ..., y_n. Наличие помех нарушает однозначное соответствие между передаваемыми и принимаемыми сообщениями. Так как помехи имеют случайный характер, то при приеме какого-либо сообщения y_j невозможно точно установить, какое сообщение было передано. Можно говорить лишь об условной вероятности $p(x_i/y_j)$,

определяющей вероятность передачи сообщения x_i при условии, что будет принято сообщение y_j.

Оценим количество информации, которое содержится в одном из принятых сообщений y_j об одном из переданных сообщений x_i.

Условная вероятность $p(x_i/y_j)$ свидетельствует о том, что имеется неопределенность в сообщении y_j относительно сообщения x_i. Эта неопределенность может быть оценена условной энтропией

$$H(x_i/y_j) = -\log_2 p(x_i/y_j). \qquad (4.17)$$

Таким образом, вследствие воздействия помех начальная априорная энтропия сообщения x_i, определяемая количественно выражением $H(x_i) = -\log_2 p(x_i)$, снимается при получении сообщения y_j не полностью, а лишь уменьшается до значения $H(x_i/y_j)$. Количество получаемой информации в этом случае будет равно снятой части неопределенности

$$I(y_j, x_i) = H(x_i) - H(x_i/y_j) = \log_2 \frac{p(x_i/y_j)}{p(x_i)}. \qquad (4.18)$$

Формула (4.18) выражает количество информации, которое содержится в принятом сообщении y_j относительно переданного x_i. Это количество информации принято называть частным количеством информации, содержащимся в сообщении y_j относительно сообщения x_i.

Среднее количество информации о всех x_i, содержащееся в одном принятом сообщении y_j, можно получить путем усреднения по всем x_i

$$I(y_j, X) = \sum_{i=1}^{n} p(x_i/y_j)\, I(y_j/x_i) = \sum_{i=1}^{n} p(x_i/y_j) \log_2 \frac{p(x_i/y_j)}{p(x_i)}. \qquad (4.19)$$

Формула (4.19) выражает частное количество информации, содержащееся в принятом сообщении y_j относительно всей совокупности переданных сообщений X.

Наконец, для того чтобы определить количество информации, содержащееся во всей совокупности принятых сообщений Y относительно всей совокупности переданных сообщений X, необходимо осуществить усреднение по всем y_j

$$I(Y, X) = \sum_{j=1}^{n} p(y_j)\, I(y_j, X) = \sum_{j=1}^{n} p(y_j) \times$$

$$\times \sum_{i=1}^{n} p(x_i/y_j) \log_2 \frac{p(x_i/y_j)}{p(x_i)} = \sum_{j=1}^{n} \sum_{i=1}^{n} p(y_j)\, p(x_i/y_j) \log_2 \frac{p(x_i/y_j)}{p(x_i)}. \qquad (4.20)$$

Используя равенство

$$p(y_j)\, p(x_i/y_j) = p(x_i, y_j),$$

получим

$$I(Y,\ X)=\sum_{i=1}^{n}\sum_{j=1}^{n}p(x_i,\ y_j)\log_2\frac{p(x_i,\ y_j)}{p(x_i)\,p(y_j)}\ . \qquad (4.21)$$

Формулы (4.20) и (4.21) определяют среднее количество информации, содержащееся в Y относительно X.

Формулу (4.21) можно привести к виду

$$I(Y,\ X)=\sum_{i=1}^{n}\sum_{j=1}^{n}p(x_i,\ y_j)\log_2 p(x_i/y_j)-$$
$$-\sum_{i=1}^{n}\sum_{j=1}^{n}p(x_i,\ y_j)\log_2 p(x_i)=-H(X/Y)-$$
$$-\sum_{i=1}^{n}\sum_{j=1}^{n}p(x_i)\,p(y_j/x_i)\log_2 p(x_i). \qquad (4.22)$$

Произведя преобразование во втором слагаемом правой части выражения (4.22) с учетом равенства $\sum_{j=1}^{n=1}p(y_j/x_i)=1$, получим

$$I(Y,\ X)=-H(X/Y)-\sum_{i=1}^{n}p(x_i)\log_2 p(x_i)=H(X)-H(X/Y). \qquad (4.23)$$

Таким образом, среднее количество информации, получаемое при неполной достоверности сообщений, равно разности безусловной энтропии $H(X)$, характеризующей начальную (априорную) неопределенность сообщений, и условной энтропии $H(X/Y)$, характеризующей остаточную (апостериорную) неопределенность сообщений.

Используя свойство условной энтропии

$$H(X/Y)=H(X,\ Y)-H(Y),$$

выражение (4.23) можно привести к виду

$$I(Y,\ X)=H(X)+H(Y)-H(X,\ Y). \qquad (4.24)$$

Следовательно, количество передаваемой информации может быть выражено через сумму энтропий передаваемого X и принимаемого Y сообщений за вычетом совместной энтропии $H(X,\ Y)$.

Так как

$$H(X,\ Y)=H(Y,\ X),$$

то

$$H(X)+H(Y/X)=H(Y)+H(X/Y),$$

или

$$H(X)-H(X/Y)=H(Y)-H(Y/X).$$

Из последнего равенства следует, что

$$I(Y,\ X) = I(X,\ Y), \tag{4.25}$$

т. е. количество информации, которое содержится в сообщении Y относительно сообщения X, равно количеству информации, содержащемуся в X относительно Y. Поэтому $I(Y,\ X)$ и $I(X,\ Y)$ называют также полной взаимной информацией.

2. Непрерывные сообщения

Для определения количества информации при непрерывных сообщениях воспользуемся в качестве исходного выражением (4.23), определяющим количество информации при дискретных сообщениях. В этом выражении начальная и остаточная энтропии сообщений соответственно

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i),$$

$$H(X/Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_i,\ y_j)\log_2 p(x_i/y_j). \tag{4.26}$$

Для перехода к непрерывным случайным величинам выразим вероятности через функции распределения плотности вероятности

$$p(x_i) = w(x_i)\Delta x_i,$$
$$p(y_j) = w(y_j)\Delta y_j;$$
$$p(x_i,\ y_j) = w(x_i,\ y_j)\Delta x_i\Delta y_j;$$
$$p(x_i/y_j) = w(x_i/y_j)\Delta x_i,$$

где Δx_i и Δy_j — элементарные участки, на которые разбиты шкалы уровней случайных величин x и y, $w(x_i)$, $w(y_j)$, $w(x_i,\ y_j)$, $w(x_i/y_j)$ — значения функций распределения при аргументах $X = x_i$, $Y = y_j$.

Тогда, произведя в (4.26) соответствующие замены, получим

$$I(Y,\ X) = -\sum_{i=1}^{n} w(x_i)\Delta x_i\log_2[w(x_i)\Delta x_i] +$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{n} w(x_i,\ y_j)\Delta x_i\Delta y_j\log_2[w(x_i/y_j)\Delta x_i].$$

Осуществим предельный переход при $\Delta x_i \to 0$ и $\Delta y_j \to 0$.

Предел выражения для начальной энтропии $H(X)$ при $\Delta x_i \to 0$

$$\lim_{\Delta x_i\to 0} H(X) = -\int_{-\infty}^{\infty} w(X)\log_2 w(X)\,dX - \lim_{\Delta x_i\to 0}\log_2\Delta x_i.$$

Определим предел для остаточной энтропии

$$\lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0}} \left\{\sum_{i=1}^{n} \sum_{j=1}^{n} w(x_i, y_j) \Delta x_i \Delta y_j \log_2 [w(x_i/y_j) \Delta x_i]\right\} =$$

$$= \lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0}} \left\{\sum_{i=1}^{n} \sum_{j=1}^{n} w(x_i, y_j) \Delta x_i \Delta y_j \log_2 w(x_i/y_j)\right\} +$$

$$+ \lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0}} \left\{\sum_{i=1}^{n} \sum_{j=1}^{n} w(x_i, y_j) \Delta x_i \Delta y_j \log_2 \Delta x_i\right\} =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(X/Y)\, dX\, dY +$$

$$+ \lim_{\substack{\Delta x_i \to 0 \\ \Delta y_j \to 0}} \left\{\log_2 \Delta x_i \sum_{i=1}^{n} \sum_{j=1}^{n} w(x_i, y_j) \Delta x_i \Delta y_j\right\} =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(X/Y)\, dX\, dY + \lim_{\Delta x_i \to 0} \log_2 \Delta x_i,$$

так как

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w(x_i, y_j) \Delta x_i \Delta y_j = 1.$$

Таким образом, после осуществления предельных переходов получаем выражение для количества информации при непрерывных сообщениях

$$I(Y, X) = h(X) - h(X/Y) = -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX +$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(X/Y)\, dX\, dY. \tag{4.27}$$

Следовательно, и при непрерывных сообщениях количество передаваемой информации определяется разностью начальной и остаточной энтропии сообщения.

Если учесть, что

$$w(X/Y) = \frac{w(X, Y)}{w(Y)} \text{ и } w(X) = \int_{-\infty}^{\infty} w(X, Y)\, dY,$$

то выражение (4.27) можно представить в ином виде:

$$I(Y, X) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X, Y) \log_2 w(X)\, dX\, dY +$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X,\ Y)\log_2\frac{w(X,\ Y)}{w(Y)}\,dX\,dY =$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X,\ Y)\log_2\frac{w(X,\ Y)}{w(X)\,w(Y)}\,dX\,dY. \qquad (4.28)$$

В отличие от дифференциальной энтропии количество информации не зависит от масштаба непрерывных сообщений, если масштабы X и Y одинаковы.

В заключение рассмотрим два предельных случая передачи сообщений.

Полная статистическая зависимость передаваемых X и принимаемых Y сообщений.

Практически это имеет место при незначительном уровне помех или при полном отсутствии помех.

В этом случае условная энтропия

$$H(X/Y) = 0.$$

Следовательно, количество информации, содержащееся в Y относительно X, равно энтропии передаваемых сообщений.

Полная статистическая независимость сообщений X и Y. Это имеет место при высоком уровне помех, когда помехи полностью подавляют полезный сигнал.

В этом случае условная энтропия равна начальной и количество информации, содержащееся в Y относительно X

$$I(Y,\ X) = 0,$$

т. е. сообщение Y не содержит никакой информации о сообщении X.

Из рассмотрения частных случаев следует, что информация, содержащаяся в Y относительно X, не превосходит энтропии X. Максимальное количество информации в случае абсолютно достоверной передачи сообщений

$$I(Y,\ X)_{\text{макс}} = H(X). \qquad (4.29)$$

Выражение (4.29) справедливо лишь для дискретных сообщений.

При абсолютно достоверной передаче непрерывных сообщений количество информации не равно ни энтропии $H(X)$, ни дифференциальной энтропии $h(X)$ сообщения. Так как непрерывные сообщения воспроизводятся с ограниченной точностью, то количество информации зависит не только от статистики сообщения X, но и от способа его воспроизведения. При этом количество информации, содержащееся в отсчетах X^* относительно сообщения X, определится разностью дифференциальных энтропий

$$I(X^*,\ X) = h(X) - h(X/X^*), \qquad (4.30)$$

где $h(X/X^*)$ — условная дифференциальная энтропия, характери-

зующая потерю информации за счет ограниченной точности воспроизведения сообщения.

Выражение (4.30) можно привести к виду

$$I(X^*, X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X, X^*) \log_2 \frac{w(X, X^*)}{w(X)\, w(X^*)}\, dX\, dY.$$

Возможны случаи, когда задаются требования к верности воспроизведения X. В качестве критерия верности может быть использовано допустимое значение среднеквадратичного отклонения X^* от X, т. е.

$$\sigma^* = \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X - X^*)\, w(X, X^*)\, dX\, dX^*}.$$

Варьируя условной плотностью вероятности $w(X/X^*) = \frac{w(X, X^*)}{w(X^*)}$, можно изменить количество информации $I(X^*, X)$. Очевидно, что наиболее выгодным будет то значение функции $w(X/X^*)$, при котором $I(X^*, X)$ имеет наименьшее значение, ибо при этом обеспечивается заданное требование к верности воспроизведения при получении минимального количества информации. Наименьшее значение $I(X^*, X)$, при котором удовлетворяется это требование, называют эпсилон-энтропией [60]. Согласно определению, эпсилон-энтропия равна

$$H_\varepsilon(X) = \text{мин}\, I(X^*, X) = h(X) - \text{макс}\, h(X/X^*)$$

при

$$\sigma^* \leqslant \varepsilon_0,$$

где ε_0 — допустимое значение ошибки воспроизведения, эпсилон-энтропия определяет информационную емкость источника непрерывных сообщений при заданном критерии верности воспроизведения [9].

В общем случае эпсилон-энтропия определяется как минимальное количество информации, определяемое двумя ситуациями (случайными величинами), одна из которых задана и мера различия между которыми определяется некоторым условием (ε) [2].

4.6. Энтропия и количество информации при статистической зависимости элементов сообщений

При определении энтропии и количества передаваемой информации предполагалось, что элементы сообщений статистически независимы. Однако в реальных условиях независимость элементов сообщений — явление довольно редкое. Например, при передаче

русского текста вероятности появления отдельных букв зависят от того, какие буквы им предшествовали. Например, если передана буква «*п*», вероятность того, что следующей буквой может быть «*а*» гораздо больше, чем вероятность появления буквы «*р*». После буквы «ъ» никогда не ожидается появление буквы «*н*» и т. п. Такая зависимость между элементами образовалась исторически в процессе длительного формирования современного русского языка.

Очевидно, что при определении энтропии и информации в сообщениях, элементы которых коррелированы, нельзя ограничиваться, как это мы делали до сих пор, только безусловными вероятностями элементов сообщений, необходимо учитывать и условные вероятности появления элементов.

Будем полагать, что передается конечное число сообщений x_1, x_2, x_3, ..., x_g, x_n, x_i, x_j, ..., x_{n-1}, x_n. Коррелятивные связи между элементами сообщений могут распространяться на различные группы элементов. Если элементы сообщений независимы, то условная вероятность передачи элемента x_j будет равна безусловной

$$p(x_j/x_1,\ x_n,\ x_g,\ \ldots,\ x_2,\ x_1) = p(x_j).$$

Если имеется коррелятивная связь только между двумя соседними элементами, то вероятность передачи любого элемента сообщения будет зависеть лишь от того, каков был предшествующий символ, т. е. условная вероятность передачи элемента x_j будет равна $p(x_j/x_i)$. В этом случае элементы сообщений составляют простую односвязную цепь Маркова.

Если коррелятивные связи охватывают три элемента сообщений, то последние представляют двухсвязную цепь Маркова и условная вероятность передачи элемента x_j будет равна $p(x_j/x_i x_n)$.

Большинство сообщений в реальных условиях представляют собой эргодическую последовательность, у которой коррелятивные связи распространяются на конечное число элементов. При достаточной длине такой последовательности с достаточной точностью могут быть определены вероятности и условные вероятности появления отдельных сообщений. Язык является типичным примером такой последовательности. В любой книге на данном языке (кроме узкоспециальных) частота повторения отдельных букв и их различных сочетаний будет постоянной, независимой от содержания книги.

Пусть сообщения составляют простую, т. е. односвязную, цепь Маркова. В этом случае энтропия элемента x_j будет определяться условной вероятностью $p(x_j/x_i)$.

Для данного фиксированного x_i энтропия сообщений будет определяться частной условной энтропией

$$H(X/x_i) = -\sum_{j=1}^{n} p(x_j/x_i) \log_2 p(x_j/x_i).$$

Произведя усреднение по всем x_i, получим выражение для средней энтропии сообщения

$$H(X)=\sum_{i=1}^{n} p(x_i)\,H(X/x_i)=-\sum_{i=1}^{n} p(x_i)\sum_{j=1}^{n} p(x_j/x_i)\log_2 p(x_j/x_i)=$$
$$=-\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_i,\ x_j)\log_2 p(x_j/x_i). \qquad (4.31)$$

В случае наличия коррелятивных связей между тремя элементами энтропия сообщений будет равна

$$H(X)=-\sum_{n=1}^{n}\sum_{i=1}^{n} p(x_i,\ x_n)\sum_{j=1}^{n} p(x_j/x_i,\ x_n)\log_2 p(x_j/x_i,\ x_n)=$$
$$=\sum_{n=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_j,\ x_i,\ x_n)\log_2 p(x_j/x_i,\ x_n). \qquad (4.32)$$

Если коррелятивными связями охвачено большее число элементов, то энтропия определится аналогично.

При наличии коррелятивных связей между элементами энтропия сообщений, а следовательно, и количество передаваемой информации уменьшаются, причем это уменьшение будет тем интенсивнее, чем сильнее коррелятивные связи и чем большее число элементов будет охвачено этими связями.

4.7. Избыточность сообщений

Как уже отмечалось, средняя энтропия сообщений при одинаковом количестве элементов может быть различной в зависимости от статистических характеристик сообщений. Энтропия максимальна и определяется выражением (4.5), если элементы сообщений равновероятны и взаимно независимы. Если поступление элементов сообщений не равновероятно, то энтропия уменьшается и определяется выражением (4.4).

Еще меньшей будет энтропия при наличии коррелятивных связей между элементами сообщений. Для случаев, когда коррелятивные связи охватывают два и три элемента, энтропия определяется выражениями (4.29) и (4.30). Сообщения, энтропия которых максимальна, являются оптимальными с точки зрения наибольшего количества передаваемой информации.

Мерой количественной оценки того, насколько данное реальное сообщение по своей энтропии отличается от соответствующего ему оптимального сообщения, является коэффициент сжатия

$$\mu=\frac{H(X)}{H(X)_{\text{макс}}},$$

где $H(X)$ — энтропия реального сообщения; $H(X)_{\text{макс}}$ — энтропия соответствующего ему оптимального сообщения.

Если неоптимальное и оптимальное сообщения характеризуются одинаковой общей энтропией, то справедливо равенство

$$nH(X) = n'H(X)_{\text{макс}},$$

где n — число элементов неоптимального сообщения; n' — число элементов соответствующего оптимального сообщения.

Так как средняя на элемент энтропия оптимального сообщения максимальна, то число элементов неоптимального сообщения всегда будет больше числа элементов соответствующего ему оптимального сообщения.

Коэффициент сжатия можно выразить через количество элементов сообщений

$$\mu = \frac{H(X)}{H(X)_{\text{макс}}} = \frac{n'}{n}.$$

Таким образом, реальные сообщения при одинаковой информативности обладают определенной избыточностью в элементах по сравнению с оптимальными сообщениями.

Мерой количественной оценки избыточности является коэффициент избыточности

$$K_{\text{и}} = \frac{n - n'}{n} = \frac{H(X)_{\text{макс}} - H(X)}{H(X)_{\text{макс}}} = 1 - \mu. \tag{4.33}$$

Избыточность приводит к увеличению времени передачи сообщений, излишней загрузке канала связи. Однако не всегда нужно стремиться к тому, чтобы избыточность $K_{\text{и}} = 0$. Некоторая избыточность бывает полезной для обеспечения требуемой надежности систем, повышения помехоустойчивости передачи сообщений. Например, иногда практикуется повторная передача одного и того же сообщения с целью повышения достоверности передачи в условиях воздействия помех.

Имеется определенная избыточность и в русском языке. Оценка статистики русского языка показывает следующее.

Средняя вероятность повторения отдельных букв в русском языке иллюстрируется табл. 2.

Таблица 2

Буквы	о	е	а	и	т	н	с	р
Вероятность повторения	0,11	0,087	0,075	0,075	0,065	0,065	0,065	0,048

Буквы	в	л	к	ц	щ	э	ф
Вероятность повторения	0,046	0,042	0,034	0,005	0,004	0,003	0,002

Русский алфавит содержит 31 букву (при условии, если не различать буквы «е» и «ё», а также мягкий и твердый знаки). С учетом пробела между буквами русский язык, таким образом, обладает 32 символами.

При условии равновероятности и независимости символов средняя энтропия на символ будет максимальна

$$H(X)_{\text{макс}} = \log_2 32 = 5 \, \frac{\text{дв. ед.}}{\text{симв.}}.$$

Если учесть различную вероятность символов, то

$$H_1(X) = 4{,}39 \, \frac{\text{дв. ед.}}{\text{симв.}}.$$

С учетом корреляции между двумя символами энтропия уменьшается до значения

$$H_2(X) = 3{,}52 \, \frac{\text{дв. ед.}}{\text{симв.}},$$

между тремя символами — до значения

$$H_3(X) = 3{,}05 \, \frac{\text{дв. ед.}}{\text{симв.}},$$

между восемью символами — до значения

$$H_8(X) = 2 \, \frac{\text{дв. ед.}}{\text{симв.}}$$

и дальше остается неизменной.

Следовательно, избыточность русского языка составляет

$$K_{\text{и}} = \frac{H(X)_{\text{макс}} - H(X)}{H(X)_{\text{макс}}} = \frac{5-2}{5} = 0{,}6.$$

Следует заметить, что у всех европейских языков избыточность примерно одинакова. Избыточность разговорных языков сформировалась в результате очень длительной общественной практики и позволяет восстанавливать целые слова и фразы при их искажениях под воздействием различных мешающих факторов.

4.8. Преимущества и недостатки статистической меры количества информации

Такой подход к определению количественной меры информации, когда в основу определения количества информации положены статистические (вероятностные) характеристики событий, имеет определенные достоинства и недостатки.

Основным достоинством статистической меры количества информации является ее универсальность. Информация измеряется в од-

них и тех же единицах, независимо от ее физической природы и содержания. В связи с этим рассмотренная мера количества информации является во многих случаях весьма удобной при анализе и синтезе сложных информационных систем и наиболее рациональной при оптимизации характеристик систем. Важным преимуществом статистической меры количества информации является ее объективность. Статистическая оценка событий не зависит от психологических факторов, ибо устанавливается на основе экспериментальных данных.

Вместе с тем то обстоятельство, что в основу измерения количества информации положены лишь статистические характеристики событий, ограничивает возможности данной меры информации.

В ряде случаев важно оценить смысловое содержание (семантику), а также ценность или целесообразность информации. Статистическая мера количества информации не в состоянии обеспечить решение данных задач.

В последнее время делаются попытки расширения области применения теории информации за счет введения семантических мер информации, мер содержательности, целесообразности информации и пр. Более подробное изложение этого вопроса смотрим в [28, 29, 30].

4.9. Примеры

Пример 4.1. Определить энтропию системы, которая описывается дискретной случайной величиной x со следующим рядом распределения

$$p(x_1) = p(x_2) = p(x_3) = p(x_4) = 0{,}01, \quad p(x_5) = 0{,}96.$$

Решение. Используя выражение (4.2) для определения энтропии дискретного распределения, получим

$$H(X) = -\sum_{i=1}^{5} p(x_i) \log_2 p(x_i) = 4 \cdot 0{,}01 \log_2 0{,}01 - 0{,}96 \log_2 0{,}96 =$$

$$= 0{,}322 \ \frac{\text{дв. ед.}}{\text{симв.}}$$

Пример 4.2. Определить энтропию сообщения из пяти букв, если общее число букв в алфавите равно 32 и все сообщения равновероятны.

Решение. Общее число пятибуквенных сообщений

$$n = 32^5.$$

Используя выражение (4.5) для определения энтропии равновероятных сообщений, получим

$$H(X) = \log_2 n = 5 \log_2 32 = 25 \ \frac{\text{дв. ед.}}{\text{симв.}}$$

Пример 4.3. Определить дифференциальную энтропию непрерывного сообщения, распределенного по нормальному закону

$$w(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a_x)^2}{2\sigma^2}\right\}.$$

Решение. Используя выражение (4.9) для определения искомой дифференциальной энтропии, получим

$$h(X) = -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a_x)^2}{2\sigma^2}\right] \times$$

$$\times \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a_x)^2}{2\sigma^2}\right\} dX = \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a_x)^2}{2\sigma^2}\right\} dX +$$

$$+ \frac{\log_2 e}{2\sigma^2} \int_{-\infty}^{\infty} (x-a_x)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a_x)^2}{2\sigma^2}\right\} dX.$$

Интеграл первого слагаемого полученного выражения равен единице, так как под интегралом имеется функция $w(X)$, а $\int_{-\infty}^{\infty} w(X)\, dX = 1$. Второе слагаемое полученного выражения

$$\frac{\log_2 e}{2\sigma^2} \int_{-\infty}^{\infty} (x-a_x)^2 w(X)\, dX = \frac{\log_2 e}{2\sigma^2} \sigma^2 = \frac{1}{2} \log_2 e.$$

Таким образом,

$$h(X) = \log_2 \sqrt{2\pi\sigma^2} + \frac{1}{2} \log_2 e = \log_2 \sqrt{2\pi e\sigma^2} \text{ дв. ед.}$$

Пример 4.4. Измеряемая величина изменяется в пределах от x_0 до $x_0 + a$ и распределена по закону равной вероятности. Найти дифференциальную энтропию данной величины.

Решение. Закон равной вероятности можно аналитически представить в виде

$$w(X) = \begin{cases} \frac{1}{a} \text{ при } x_0 \leqslant X \leqslant x_0 + a, \\ 0 \text{ при } X < x_0 \text{ и } X > x_0 + a. \end{cases}$$

Тогда, используя (4.9), получим искомую энтропию

$$h(X) = -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX = -\frac{1}{a} \int_{x_0}^{x_0+a} \log_2 \frac{1}{a} dX = \log_2 a \text{ дв. ед.}$$

Пример 4.5. Ансамбли событий $X(x_1, x_2, x_3)$ и $Y(y_1, y_2)$ объединены. Вероятности совместных событий $p(X, Y)$ приведены в табл. 3.

Таблица 3

y_j \ x_i	x_1	x_2	x_3
y_1	0,1	0,2	0,3
y_2	0,25	0	0,15

Определить:

а) энтропию ансамблей X и Y;

б) энтропию объединенного ансамбля X, Y;

в) условные энтропии ансамблей.

Решение. Из очевидных равенств

$$\sum_{j=1}^{n} p(x_i, y_j) = p(x_i); \quad \sum_{i=1}^{n} p(x_i, y_j) = p(y_j)$$

получаем

$$p(x_1) = p(x_1, y_1) + p(x_1, y_2) = 0,1 + 0,25 = 0,35;$$
$$p(x_2) = p(x_2, y_1) + p(x_2, y_2) = 0,2;$$
$$p(x_3) = p(x_3, y_1) + p(x_3, y_2) = 0,3 + 0,15 = 0,45;$$
$$p(y_1) = p(x_1, y_1) + p(x_2, y_1) + p(x_3, y_1) = 0,1 + 0,2 + 0,3 = 0,6;$$
$$p(y_2) = p(x_1, y_2) + p(x_1, y_2) + p(x_3, y_2) = 0,25 + 0,15 = 0,4.$$

Тогда энтропии событий

$$H(X) = -\sum_{i=1}^{3} p(x_i) \log_2 p(x_i) = -0,35 \log_2 0,35 - 0,2 \log_2 0,2 -$$
$$- 0,45 \log_2 0,45 = 1,512 \, \frac{\text{дв. ед.}}{\text{сообщ.}};$$

$$H(Y) = -\sum_{j=1}^{2} p(y_j) \log_2 p(y_j) = -0,6 \log_2 0,6 - 0,4 \log_2 0,4 =$$
$$= 0,971 \, \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Энтропию объединенного ансамбля находим, используя (4.7)

$$H(X, Y) = -\sum_{i=1}^{3} \sum_{j=1}^{2} p(x_i, y_j) \log_2 p(x_i, y_j) = -0,1 \log_2 0,1 -$$
$$- 0,25 \log_2 0,25 - 0,2 \log_2 0,2 - 0,3 \log_2 0,3 - 0,15 \log_2 0,15 =$$
$$= 2,228 \, \frac{\text{дв. ед.}}{\text{сообщ.}}$$

Условные энтропии удобно определить, используя свойство (4.12)

$$H(Y/X) = H(X, Y) - H(X) = 2{,}228 - 1{,}512 = 0{,}716 \ \frac{\text{дв. ед.}}{\text{сообщ.}};$$

$$H(X/Y) = H(X, Y) - H(Y) = 2{,}228 - 0{,}971 = 1{,}257 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Пример 4.6. Сигнал формируется в виде двоичного кода с вероятностями появления символов 1 и 0, равными соответственно $p(x_1) = 0{,}6$ и $p(x_0) = 0{,}4$. Появления любого из символов взаимосвязаны условными вероятностями:

$p(x_0/x_0) = 0{,}1$ — вероятность того, что после 0 будет 0;
$p(x_1/x_0) = 0{,}9$ — вероятность того, что после 0 будет 1;
$p(x_1/x_1) = 0{,}1$ — вероятность того, что после 1 будет 1;
$p(x_0/x_1) = 0{,}9$ — вероятность того, что после 1 будет 0.

Найти энтропию сигналов.

Решение. Используя (4.31), находим искомую энтропию

$$H(X) = -\sum_{i=0}^{1} p(x_i) \sum_{j=0}^{1} p(x_i/x_j) \log_2 p(x_j/x_i) =$$

$$= -(0{,}1 \log_2 0{,}1 + 0{,}9 \log_2 0{,}9)(0{,}6 + 0{,}4) = 0{,}467 \ \frac{\text{дв. ед.}}{\text{эл.}}.$$

Пример 4.7 [101]. Вероятности появления сообщений x_1, x_2, x_3 и x_4 равны $p(x_1) = \frac{1}{2}$; $p(x_2) = \frac{1}{4}$; $p(x_3) = p(x_4) = \frac{1}{8}$.

Между сообщениями имеются корреляционные связи, описанные табл. 4. Найти энтропию сообщений.

Таблица 4

(x_i, x_j)	$p(x_i, x_j)$	$p(x_i/x_j)$	x_i, x_j	$p(x_i, x_j)$	$p(x_i/x_j)$
x_1x_1	13/32	13/16	x_3x_1	0	0
x_1x_2	3/32	3/16	x_3x_2	0	0
x_1x_3	0	0	x_3x_3	0	0
x_1x_4	0	0	x_3x_4	1/8	1
x_2x_1	1/32	1/8	x_4x_1	1/16	1/2
x_2x_2	1/8	1/2	x_4x_2	1/32	1/4
x_2x_3	3/32	3/8	x_4x_3	1/32	1/4
x_2x_4	0	0	x_4x_4	0	0

Решение. Используя (4.31), получим искомую энтропию

$$H(X) = -\sum_{i=1}^{4} \sum_{j=1}^{4} p(x_i, x_j) \log_2 p(x_i/x_j) = 13/32 \log_2 13/16 -$$

$$- 3/32 \log_2 3/16 - 1/32 \log_2 1/8 - 1/8 \log_2 1/2 - 3/32 \log_2 3/8 -$$

$$- 1/16 \log_2 1/2 - 2/32 \log_2 1/4 = 0{,}886 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Пример 4.8 [91]. Задающее воздействие X в виде электрического напряжения имеет 17 независимых дискретных значений с шагом квантования Δ, вероятности появления которых распределены по двухстороннему экспоненциальному закону с плотностью распределения $w(X) = \frac{1}{2a}\exp\left\{-\frac{|x|}{a}\right\}$ (рис. 4.4), причем $a = 0{,}53$, $\Delta = 0{,}263$.

Определить энтропию дискретного сигнала X.

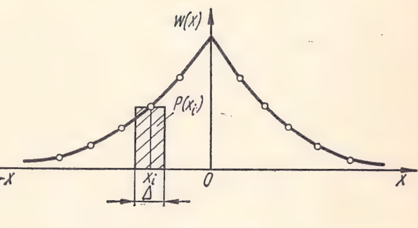


Рис. 4.4

Решение. Вероятности появления уровней сигнала (рис. 4.4) определим по приближенной формуле

$$p(x_i) = w(x_i)\,\Delta = 0{,}2\exp\left\{-\frac{|x|}{0{,}5}\right\}.$$

Вычисления сведены в табл. 5.

Таблица 5

x_i	0	1	2	3	4	5	6	7	8
x_i	0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6
$w(x_i)$	1,00	0,67	0,45	0,30	0,20	0,14	0,091	0,061	0,041
$p(x_i)$	0,200	0,134	0,090	0,060	0,040	0,028	0,018	0,012	0,008

Так как функция $w(X)$ симметрична, то вероятности для значений x_i и $-x_i$ одинаковы. Тогда искомая энтропия будет равна

$$H(X) = -p(x_0)\log_2 p(x_0) - 2\sum_{i=1}^{8} p(x_i)\log_2 p(x_i) = 0{,}464\ \frac{\text{дв. ед.}}{\text{отсчет}}.$$

Пример 4.9. Найти совместную энтропию взаимосвязанных сообщений X и Y и условную энтропию $H(X/Y)$, если двумерная функция плотности распределения подчинена нормальному закону

$$w(X,\ Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp\left[-\frac{1}{(2-r^2)}\left(\frac{x^2}{\sigma_x^2} - \frac{2r\,xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right)\right],$$

где r — коэффициент корреляции значений X и Y.

Решение. Совместная дифференциальная энтропия X и Y определяется соотношением

$$h(X,\ Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X,\ Y)\log_2 w(X,\ Y)\,dX\,dY,$$

где

$$\log_2 w(X,\ Y) = -\log_2\left(2\pi\sigma_x\sigma_y\sqrt{1-r^2}\right) - \frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_x}\frac{2r\,xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right)\log_2 e.$$

Подставляя значение $\log_2 w(X, Y)$ в выражение для $h(X, Y)$, получим

$$h(X, Y) = \log_2 \left(2\pi\sigma_x\sigma_y \sqrt{1-r^2}\right) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X, Y)\, dX\, dY +$$

$$+ \frac{\log_2 e}{2\sigma_x^2(1-r^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X^2 w(X, Y)\, dX\, dY +$$

$$+ \frac{\log_2 e}{2\sigma_y^2(1-r^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y^2 w(X, Y)\, dX\, dY -$$

$$- \frac{r\log_2 e}{\sigma_x\sigma_y(1-r^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY w(X, Y)\, dX\, dY.$$

Так как

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(X, Y)\, dX\, dY = 1; \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY w(X, Y)\, dX\, dY = r\,\sigma_x\sigma_y;$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X^2 w(X, Y)\, dX\, dY = \sigma_x^2; \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y^2 w(X, Y)\, dX\, dY = \sigma_y^2,$$

то окончательно получим

$$h(X, Y) = -\log_2 \left(2\pi\sigma_x\sigma_y\sqrt{1-r^2}\right) + \frac{1}{1-r^2}\log_2 e - \frac{r^2}{1-r^2}\log_2 e =$$

$$= \log_2\left(2\pi e\sigma_x\sigma_y\sqrt{1-r^2}\right) \frac{\text{дв. ед.}}{\text{отсчет.}}.$$

Применяя формулу (4.11) и учитывая, что

$$h(Y) = \log_2\left(\sqrt{2\pi e}\,\sigma_y\right),$$

получим значение условной энтропии

$$h(X/Y) = h(X, Y) - h(Y) = \log_2\left[\sqrt{2\pi e(1-r^2)}\,\sigma_x\right] \frac{\text{дв. ед.}}{\text{отсчет.}}.$$

Пример 4.10 [27]. Контролируемый параметр X может принимать два значения x_0 и x_1 с равными вероятностями $p(x_0) = p(x_1)$. Вследствие ограниченной точности системы контроля будут иметь место ошибки контроля, т. е. вместо x_0 может быть зафиксировано x_1 и, наоборот, вместо x_1 зафиксировано x_0. Условные вероятности таких событий равны 0,01.

Определить количество информации, получаемое при контроле.

Решение. Введем следующие обозначения результатов контроля: y_0 — система контроля указывает, что параметр X имеет значение x_0; y_1 — система контроля указывает, что параметр X имеет значение x_1.

Тогда условные вероятности ошибочного контроля будут соответственно равны $p(x_1/x_0) = 0{,}01$ и $p(x_0/x_1) = 0{,}01$.

На основании (4.23) количество получаемой информации при неполной достоверности контроля будет равно разности начальной и остаточной энтропии

$$I(Y,\ X) = H(X) - H(X/Y) = -\sum_{i=1}^{1} p(x_i)\log_2 p(x_i) + \\ + \sum_{j=0}^{1} p(y_j) \sum_{i=0}^{1} p(x_i/y_j)\log_2 p(x_i/y_j).$$

Начальная энтропия

$$H(X) = -p(x_0)\log_2 p(x_0) - p(x_1)\log_2 p(x_1) = \\ = 2 \cdot 0{,}5\log_2 0{,}5 = 1\ \frac{\text{дв. ед.}}{\text{сообщ.}}$$

Для определения условной энтропии $H(X/Y)$ необходимо знать вероятности вида $p(y_j)$ и $p(x_i/y_j)$.

Вероятности событий $p(y_j)$ вычисляем по формуле полной вероятности

$$p(y_j) = \sum_i p(x_i)\,p(y_j/x_i).$$

Тогда

$$p(y_0) = p(x_0)\,p(y_0/x_0) + p(x_1)\,p(y_0/x_1) = p(x_0)\,[1 - p(y_1/x_0)] + \\ + p(x_1)\,p(y_0/x_1) = 0{,}5\,(1 - 0{,}01) + 0{,}5 \cdot 0{,}01 = 0{,}5;$$

$$p(y_1) = p(x_0)\,p(y_1/x_0) + p(x_1)\,p(y_1/x_1) = p(x_0)\,p(y_1/x_0) + \\ + p(x_i)\,[1 - p(y_0/x_1)] = 0{,}5.$$

Условные вероятности $p(x_i/y_j)$ вычисляем по теореме Байеса

$$p(x_i/y_j) = \frac{p(x_i)\,p(y_j/x_i)}{p(y_j)}.$$

Тогда

$$p(x_0/y_1) = \frac{p(x_0)\,p(y_1/x_0)}{p(y_1)} = \frac{0{,}5 \cdot 0{,}01}{0{,}5} = 0{,}01;$$

$$p(x_1/y_0) = \frac{p(x_1)\,p(y_0/x_1)}{p(y_0)} = \frac{0{,}5 \cdot 0{,}01}{0{,}5} = 0{,}01.$$

Таким образом, условная энтропия

$$H(X/Y) = -p(y_0)\,[p(x_0/y_0)\log_2 p(x_0/y_0) + p(x_1/y_0)\log_2 p(x_1/y_0)] - \\ - p(y_1)\,[p(x_0/y_1)\log_2 p(x_0/y_1) + p(x_1/y_1)\log_2 p(x_1/y_1)] = \\ -p(y_0)\,\{[1 - p(x_1/y_0)]\log_2 [1 - p(x_1/y_0)] + p(x_1/y_0)\log_2 p(x_1/y_0)\} - \\ - p(y_1)\,\{p(x_0/y_1)\log_2 p(x_0/y_1) + [1 - p(x_0/y_1)]\log_2 [1 - p(x_0/y_1)\} = \\ = 0{,}081\ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Окочательно получаем

$$I(Y, X) = 1 - 0{,}081 = 0{,}92 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Пример 4.11 [27]. Опыт α состоит в том, что мы ищем некоторое число x из 1, 2, ..., N чисел. Опыт β уточняет это число, так как он ставит условие, что искомое число должно делиться на m. Какова информация опыта β относительно опыта α?

Решение. Искомое количество информации может быть определено через разность безусловной и условной энтропий

$$H(\beta, \alpha) = H(\alpha) - H(\alpha/\beta).$$

Опыт α состоит в том, что мы из 1, 2, 3, ..., N чисел с одинаковой вероятностью можем случайно выбрать любое число. Поэтому энтропия такого опыта

$$H(\alpha) = \log_2 N.$$

Осуществляя опыт β, мы будем иметь дело с двумя исходами: число делится на m и не делится на m. Пусть из N чисел q чисел делится на m. Тогда вероятности первого и второго исходов

$$p_1 = \frac{q}{N}; \quad p_2 = \frac{N-q}{N}.$$

Частная условная энтропия, когда одно из q чисел делится на m,

$$H_1(\alpha/\beta) = \log_2 q.$$

Частная условная энтропия второго исхода опыта β

$$H_2(\alpha/\beta) = \log_2 (N - q).$$

Результирующая усредненная условная энтропия

$$H(\alpha/\beta) = p_1 \log_2 q + p_2 \log_2 (N - q) = \frac{q}{N} \log_2 q + \frac{N-q}{N} \log_2 (N-q).$$

Следовательно, получаемая информация

$$I(\beta, \alpha) = \log_2 N - \frac{q}{N} \log_2 q - \frac{N-q}{N} \log_2 (N-q).$$

Пример 4.12 [27]. Передается четыре независимых сообщения x_1, x_2, x_3 и x_4 с одинаковыми априорными вероятностями. Пусть при правильном воспроизведении информации передаваемых сообщений на приемном конце сообщению x_1 соответствует y_1, $x_2 - y_2$ и $x_3 - y_3$. Вероятности правильного и ошибочного воспроизведений сообщений соответственно равны $p = q = 0{,}5$. Определить количество информации, содержащееся в Y относительно X.

Решение. Безусловная энтропия сообщений X

$$H(X) = \log_2 4 = 2 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Условная энтропия может быть определена из соотношения

$$H(X/Y) = -\sum_{j=1}^{4} p(y_j) \sum_{i=1}^{4} p(x_i/y_j) \log_2 p(x_i/y_j),$$

где

$$p(y_1) = p(x_1)\,p(y_1/x_1) + p(x_2)\,p(y_1/x_2) + p(x_3)\,p(y_1/x_3) + \\ + p(x_4)\,p(y_1/x_4);$$

$$p(y_2) = p(x_1)\,p(y_2/x_1) + p(x_2)\,p(y_2/x_2) + p(x_3)\,p(y_2/x_3) + \\ + p(x_4)\,p(y_2/x_4);$$

$$p(y_3) = p(x_1)\,p(y_3/x_1) + p(x_2)\,p(y_3/x_2) + p(x_3)\,p(y_3/x_3) + \\ + p(x_4)\,p(y_3/y_4);$$

$$p(y_4) = p(x_1)\,p(y_4/x_1) + p(x_2)\,p(y_4/x_2) + p(x_3)\,p(y_4/x_3) + \\ + p(x_4)\,p(y_4/x_4).$$

Так как справедливо равенство

$$p(y_2/x_1) + p(y_3/x_1) + p(y_4/x_1) = q,$$

то вследствие равновероятности событий

$$p(y_2/x_1) = p(y_3/x_1) = p(y_4/x_1) = \frac{q}{3}.$$

Аналогично можно показать, что

$$p(y_1/x_2) = p(y_3/x_2) = p(y_4/x_2) = \frac{q}{3};$$

$$p(y_1/x_3) = p(x_2/x_3) = p(y_4/x_3) = \frac{q}{3};$$

$$p(y_1/x_4) = p(y_2/x_4) = p(y_3/x_4) = \frac{q}{3}.$$

С учетом того, что $p(y_1/x_1) = p(y_2/x_2) = p(y_3/x_3) = p(y_4/x_4) = p$, получим

$$p(y_1) = \frac{1}{4}\left(p + \frac{q}{3} + \frac{q}{3} + \frac{q}{3}\right) = \frac{1}{4}.$$

Аналогично определяем

$$p(y_2) = \frac{1}{4}; \quad p(y_3) = \frac{1}{4}; \quad p(y_4) = \frac{1}{4}.$$

Условные вероятности $p(x_i/y_j)$ находим на основании формулы Байеса

$$p(x_i/y_j) = \frac{p(x_i)\,p(y_j/x_i)}{p(y_j)}.$$

Тогда

$$p(x_1/y_1) = \frac{p(x_1)\,p(y_1/x_1)}{p(y_1)} = \frac{1/4p}{1/4} = p.$$

Аналогично находим

$$p(x_2/y_2) = p(x_3/y_3) = p(x_4/y_4) = p;$$

$$p(x_2/y_1) = p(x_3/y_1) = p(x_4/y_1) = p(x_2/y_1) = p(x_3/y_2) = p(x_4/y_2) =$$
$$= p(x_1/y_3) = p(x_2/y_3) = p(x_4/y_3) = p(x_1/y_4) = p(x_2/y_4) =$$
$$= p(x_3/y_4) = \frac{q}{3}.$$

Следовательно, условная энтропия

$$H(X/Y) = -p(y_1)\left[p\log_2 p + 3\frac{q}{3}\log_2\frac{q}{3}\right] -$$
$$- p(y_2)\left[p\log_2 p + 3\frac{q}{3}\log_2\frac{q}{3}\right] - p(y_3)\left[p\log_2 p + 3\frac{q}{3}\log_2\frac{q}{3}\right] -$$
$$- p(y_4)\left[p\log_2 p + 3\frac{q}{3}\log_2\frac{q}{3}\right] = -p\log_2 p - q\log_2\frac{q}{3}.$$

Таким образом, получаемая информация

$$I(Y,\ X) = 2 + p\log_2 p + q\log_2\frac{q}{3} = 2 + 1/2\log_2 1/2 +$$
$$+ \frac{1}{2}\log_2 1/6 = 0{,}21\ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Пример 4.13. Определить количество информации, содержащееся в одном замере случайной величины x, равномерно распределенной в пределах от 0 до 256, если погрешность измерения распределена по нормальному закону и среднеквадратическое значение погрешности $\sigma = 5$.

Решение. Дифференциальная энтропия случайной величины X

$$h(x) = -\int_0^{256}\frac{1}{256}\log_2\frac{1}{256}\,dx = 8\ \text{дв. ед.}$$

Остаточная дифференциальная энтропия определяется погрешностью измерения и на основании примера 4.3

$$h(\delta) = \log_2\sqrt{2\pi e\sigma} = \log_2\sqrt{2\pi e5} = 4{,}39\ \text{дв. ед.}$$

Количество информации, получаемое в результате одного замера, определяется разностью начальной и конечной энтропий

$$I(x) = h(x) - h(\delta) = 8 - 4{,}39 = 3{,}61\ \text{дв. ед.}$$

Пример 4.14 [28]. Полная шкала измерительного прибора содержит 1000 делений. Погрешность прибора составляет $\pm 1\%$ от полной шкалы и распределена по закону равной вероятности. Определить избыточность шкалы прибора.

Решение. В оптимальном случае при приведенной погрешности, равной $\pm 1\%$, достаточно иметь 50 делений шкалы. Тогда на основании (4.33) избыточность

$$K_и = \frac{1000 - 50}{1000} = 0{,}95.$$

Пример 4.18. Вырабатывается стимулирующий сигнал $x(t)$ с одномерным нормальным распределением и корреляционной функцией вида

$$K_x(\tau) = \sigma_x^2 \exp\{-a(\tau)\},$$

где

$$\sigma_x = 10В; \quad a = 50\ \frac{1}{с}.$$

Отсчет сигнала производится в $m = 100$ точках. Определить энтропию квантованного сигнала и его избыточность.

Решение. Так как энтропия квантованного сигнала может быть представлена приближенным равенством

$$H(X) = -\sum_{i=1}^{m} w(x_i)\,\Delta(x) \log_2 [w(x_i)\,\Delta x],$$

где Δx — шаг квантования, то при достаточно большом числе уровней квантования операцию суммирования можно заменить интегрированием

$$H(X) = -\int_{-\infty}^{\infty} w(X) \log_2 [w(X)\,\Delta x]\,dX =$$

$$= -\int_{-\infty}^{\infty} w(X) \log_2 w(X)\,dX - \log_2 \Delta x.$$

В случае отсутствия корреляции энтропия квантованного сигнала

$$H(X)_{макс} = \log_2 (\sqrt{2\pi e}\,\sigma_x) - \log_2 \Delta x = \log_2 \left(\sqrt{2\pi e}\,\frac{\sigma_x}{\Delta_x}\right).$$

Шаг квантования определим следующим образом:

$$\Delta_x = \frac{L_x}{m},$$

где L_x — динамический диапазон сигнала.

Для нормально распределенного сигнала можно принять

$$L_x = 6\sigma_x,$$

тогда

$$\Delta_x = \frac{6\sigma_x}{m} = \frac{60}{100} = 0{,}6\ В.$$

Следовательно,

$$H(X)_{макс} = \log_2 \left(\sqrt{2\pi e}\,\frac{10}{0{,}6}\right) = 6{,}11\ \frac{дв.\ ед.}{отсчет}.$$

Энтропия квантованного сигнала при наличии корреляции в соответствии с выражением для условной энтропии коррелированного сигнала, полученным в примере 4.9,

$$H(X) = \log_2\left(\sqrt{2\pi e(1 - r^2)}\,\sigma_x\right) - \log_2 \Delta(x) =$$
$$= \log_2\left(\sqrt{2\pi e(1 - r^2)}\,\frac{\sigma_x}{\Delta_x}\right).$$

Коэффициент корреляции r может быть получен из соотношения

$$r = \frac{K_x(\tau_0)}{\sigma^2_{(x)}},$$

где τ_0 — интервал корреляции.

На основании (3.34) интервал корреляции

$$\tau_0 = \frac{1}{K_x(0)}\int_0^\infty K_x(\tau)\,d\tau = \frac{1}{\sigma_x^2}\int_0^\infty \sigma_x^2 \exp\{-a|\tau|\}\,d\tau = \frac{1}{a} = \frac{1}{50}\ \text{c}.$$

Следовательно,

$$r = \frac{K_x(\tau_0)}{\sigma_x^2} = \exp\{-a|\tau_0|\} = \exp\{-1\} = 0{,}368.$$

Таким образом, энтропия квантованного сигнала при наличии корреляции

$$H(X) = \log_2\left(\sqrt{2\pi e(1 - 0{,}368)}\,\frac{10}{0{,}6}\right) = 6\ \frac{\text{дв. ед.}}{\text{отсчет}}.$$

Избыточность в соответствии с (4.33) будет равна

$$K_\text{и} = \frac{H(X)_\text{макс} - H(X)}{H(X)_\text{макс}} = \frac{6{,}11 - 6}{6{,}11} = 0{,}02.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Поясните сущность понятия энтропии.
2. В каких единицах измеряется энтропия?
3. Сформулируйте основные свойства энтропии.
4. Как определяется энтропия дискретной системы с равновероятными и неравновероятными состояниями?
5. Чему равна энтропия при неравновероятном и взаимозависимом распределении элементов системы?
6. В чем заключается особенность определения энтропии для непрерывных распределений?
7. Что такое условная и совместная энтропия?
8. Сформулируйте основные свойства энтропии сложных сообщений.
9. Поясните связь между энтропией и информацией.
10. Какие основные требования предъявляются к мере количества информации?
11. Как количественно оценивается информация при полной и неполной достоверности сообщений?

12. Что понимается под ε-энтропией?

13. Сформулируйте преимущества и недостатки статистической меры количества информации.

14. Поясните сущность понятий частной и полной информации.

15. Что понимается под избыточностью сообщений?

16. Что является мерой количественной оценки избыточности?

Глава V

ПЕРЕДАЧА ИНФОРМАЦИИ

5.1. Обобщенные характеристики сигналов и информационных каналов

Любую информационную систему можно подразделить на источник, преобразователь, канал передачи, накопитель и отображающее устройство информации, а информационные процессы, происходящие в этих устройствах, представить в общем случае в виде процесса передачи информации по некоторому каналу.

Обобщенная схема канала передачи информации представлена на рис. 5.1. Сообщения X источника информации *ИИ*, после кодирования и модуляции в преобразователе $П_1$, превращаются в сигнал Y, поступающий в линию связи *ЛС*. В результате действия помех ξ сигнал Z на приемной стороне отличается от Y. Приемная часть содержит преобразователь $П_2$, демодулирующий и декодирующий сигналы и перерабатывающий их в сообщения W, поступающие в приемник информации *ПИ*.

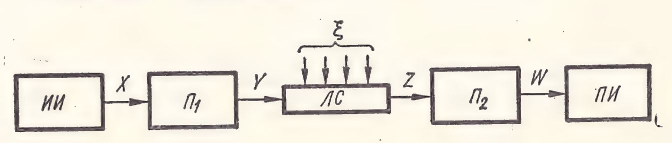


Рис. 5.1

Непосредственным переносчиком информации является сигнал, в общем случае представляющий собой случайно изменяющуюся физическую величину. На практике при рассмотрении информационных процессов удобно оперировать обобщенными показателями, характерными для множества сигналов данного вида и наиболее существенными с точки зрения передачи содержащейся в них информации. Каждый сигнал имеет определенную длительность. Длительность сигнала характеризует время передачи сообщений, продолжительность занятости информационного канала. Таким образом, вполне естественно в качестве первой обобщенной характеристики взять время передачи сигнала $T_с$. Каждый сигнал характеризуется вполне определенным частотным спектром. Теоретически ширина спектра сигнала конечной длительности неограничена. Однако

изучение спектров реальных сигналов показывает, что их спектральная плотность убывает с ростом частоты. Это позволяет при определенных условиях рассматривать сигналы как процессы с ограниченным спектром. Существуют различные критерии ограничения спектра сигнала. Одним из таких критериев являются допустимые искажения сигнала. Например, практика показала, что при передаче речевого сигнала разборчивость и качество речи практически полностью сохраняются при ширине спектра от 300 до 3400 Гц [17]. Таким образом, второй обобщенной характеристикой сигнала должна быть ширина его частотного спектра F_c. Третьей важной характеристикой сигнала является его энергетическая характеристика —

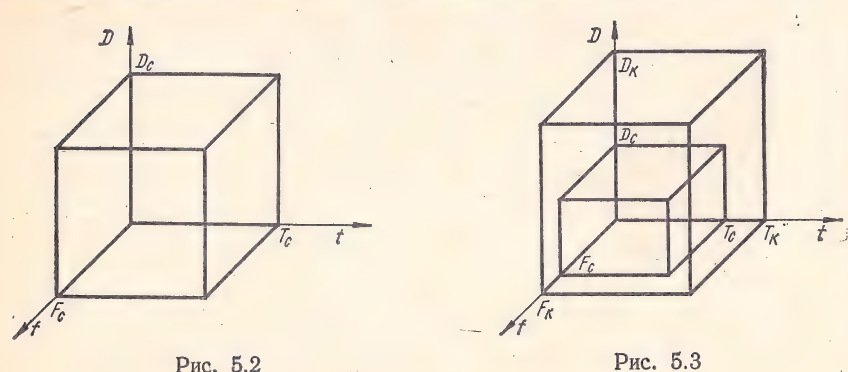


Рис. 5.2 Рис. 5.3

средняя мощность P_c. Однако поскольку при передаче на сигналы всегда воздействуют помехи, то в качестве энергетической характеристики сигнала целесообразно брать отношение средней мощности сигнала P_c к средней мощности помехи P_ξ. Обычно это отношение выражают в логарифмической мере $D = \log_2 P_c/P_\xi$ и называют динамическим диапазоном, причем при оценке информационной содержательности удобно выражать динамический диапазон через логарифм с основанием 2.

Произведение

$$V_c = T_c \cdot F_c \cdot D_c \qquad (5.1)$$

принято называть объемом сигнала. В геометрическом представлении объем сигнала имеет вид параллелепипеда с ребрами T_c, F_c и D_c (рис. 5.2).

Информационный канал можно характеризовать также тремя соответствующими параметрами: временем использования канала $T_к$, шириной полосы частот, пропускаемых каналом $F_к$, и динамическим диапазоном канала $D_к$, характеризующим его способность передавать различные уровни сигнала.

Величина

$$V_к = T_к F_к D_к \qquad (5.2)$$

называется емкостью канала.

Неискажённая передача сигналов возможна только при условии, что сигнал по своему объему «вмещается» в емкость канала (рис. 5.3).

Следовательно, общее условие согласования сигнала с каналом передачи информации определяется соотношением

$$V_c \leqslant V_к. \tag{5.3}$$

Однако соотношение (5.3) выражает необходимое, но недостаточное условие согласования сигнала с каналом. Достаточным условием является согласование по всем параметрам

$$\begin{aligned} T_c &\leqslant T_к; \\ F_c &\leqslant F_к; \\ D_c &\leqslant D_к. \end{aligned} \tag{5.4}$$

Если при выполнении условия (5.3) не обеспечивается часть условий (5.4), то можно добиться согласования трансформацией сигнала при сохранении его объема.

Например, если отсутствует согласование речевого сигнала с каналом только по частоте, т. е. имеют место соотношения

$$\begin{aligned} T_c &< T_к; \\ F_c &> F_к; \\ D_c &< D_к, \end{aligned}$$

то согласование по частоте при передаче можно достичь, записывая сигнал на магнитную ленту с одной скоростью и воспроизводя его с меньшей скоростью. Предположим, что при выполнении условий $V_c \leqslant V_к$ и $D_c \leqslant D_к$ частотный спектр сигнала в n раз шире полосы пропускания канала $F_c = nF_к$.

Для согласования сигнала с каналом можно записать сигнал на магнитную ленту со скоростью u_1, а передавать со скоростью u_2, в n раз меньшей скорости u_1. При этом длительность сигнала увеличивается в n раз и во столько же раз уменьшается ширина его спектра. Объем сигнала при этом остается неизменным. Геометрическое представление трансформации сигнала по параметрам F_c и T_c показано на рис. 5.4.

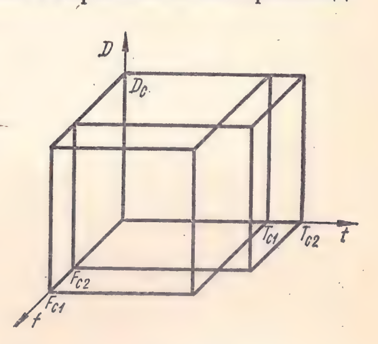


Рис. 5.4

Среднее количество информации, передаваемое по каналу в единицу времени, называется скоростью передачи информации. В общем случае скорость передачи информации зависит от длительности передачи сообщений T. При достаточно длинных сообщениях скорость передачи остается постоянной. В связи с этим аналитически

скорость передачи информации выражается следующим образом

$$\bar{I}(Z, Y) = \lim_{T \to \infty} \frac{I(Z, Y)}{T}, \quad (5.5)$$

где $I(Z, Y)$ — количество информации, передаваемое сигналом длительностью T.

Наибольшая теоретически достижимая для данного канала скорость передачи информации называется пропускной способностью канала.

Пропускная способность информационного канала равна

$$C = \text{макс}\{\bar{I}(Z, Y)\}. \quad (5.6)$$

Скорость передачи информации в общем случае зависит от статистических свойств сообщения, метода кодирования и свойств канала. Пропускная способность — это характеристика канала. Она не зависит от фактической скорости передачи информации.

С целью наиболее эффективного использования информационного канала необходимо принимать меры к тому, чтобы скорость передачи информации была как можно ближе к пропускной способности канала.

Вместе с тем скорость ввода информации в канал не должна превышать пропускную способность канала, иначе не вся информация будет передана по каналу.

Аналитически скорость ввода информации (поток информации) выражается следующим образом:

$$\bar{I}(X) = \lim_{T \to \infty} \frac{I(X)}{T}, \quad (5.7)$$

где $I(X)$ — среднее количество информации, содержащееся в сообщении на входе канала; T — длительность сообщения.

Таким образом, должно быть выполнимо условие

$$\bar{I}(X) \leqslant C. \quad (5.8)$$

Это основное условие динамического согласования источника сообщений и информационного канала.

Одним из основных вопросов в теории передачи информации является определение зависимости скорости передачи информации и пропускной способности от параметров канала и характеристик сигналов и помех. Эти вопросы были впервые глубоко исследованы Шенноном [23].

Рассмотрим три вида каналов: дискретный канал без помех, дискретный канал с помехами и непрерывный канал с помехами.

5.2. Скорость передачи информации и пропускная способность дискретного канала без помех

Под дискретным каналом передачи информации принято понимать совокупность средств, предназначенных для передачи дискретных сигналов [2].

На вход такого канала подаются дискретные сообщения X, образующие первичный алфавит $x_1, x_2, \ldots, x_n$. Последние кодируются с помощью преобразователя Π_1 (рис. 5.1) и преобразуются в кодированные сигналы Y. Для кодирования используется некоторый алфавит символов $y_1, y_2, \ldots, y_m$, а существо кодирования сводится к представлению отдельных сообщений или последовательностей сообщений определенными комбинациями символов используемого алфавита.

Скорость ввода информации

$$\bar{I}(X) = \frac{H(X)}{\bar{\tau}_x} = \bar{V}_x H(X), \tag{5.9}$$

где $H(X)$ — средняя энтропия одного сообщения; $\bar{\tau}_x$ — средняя длительность сообщения; $\bar{V}_x = \frac{1}{\bar{\tau}_x}$ — скорость выдачи символов сообщения источником.

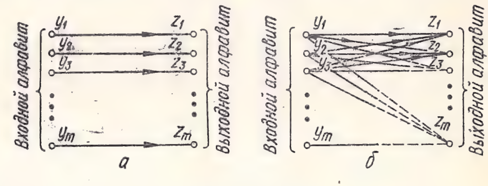


Рис. 5.5.

Под длительностью сообщения понимается интервал времени, в который появляются сообщения на выходе источника информации. Средняя длительность $\bar{\tau}_x$ при отсутствии статистических зависимостей между сообщениями определяется выражением

$$\bar{\tau}_x = \sum_{i=1}^{n} p(x_i)\, \tau_{x_i}, \tag{5.10}$$

где $p(x_i)$ и τ_{x_i} — априорная вероятность и длительность i-го сообщения; n — количество сообщений.

В канале без помех каждому определенному входному сигналу всегда будет соответствовать один и тот же сигнал на выходе канала, иными словами, входные и выходные сигналы связаны однозначной функциональной зависимостью (рис. 5.5, *а*).

В этом случае среднее количество информации, переносимое одним символом, равно энтропии символа на входе канала

$$I(Y) = H(Y)\ \frac{\text{дв. ед.}}{\text{символ}}.$$

Скорость передачи информации

$$I(Y) = \bar{U}_y \cdot H(Y)\ \frac{\text{дв. ед.}}{\text{с}}, \tag{5.11}$$

где $\bar{U}_y = \frac{1}{\bar{\tau}_y}$ — скорость передачи элементарных символов сигнала; $\bar{\tau}_y$ — средняя длительность элементарных сигналов.

Пропускная способность дискретного канала без помех

$$C = \text{макс}\,\{\bar{U}_y H(Y)\}.$$

Полагая $\overline{U}_y$ заданной, получим, что максимальная скорость передачи информации будет обеспечена при максимальном значении энтропии кодированного сигнала

$$C = \overline{U}_y \text{макс}\,\{H(Y)\} = \overline{U}_y \log_2 n, \qquad (5.12)$$

т. е. при равномерном распределении вероятностей и статистической независимости символов алфавита сигналов.

Таким образом, скорость передачи информации может быть максимальной при условии, если статистические характеристики источника сообщений определенным образом согласованы со свойствами информационного канала. Для каждого источника сообщений это согласование может быть достигнуто специальным выбором способа кодирования сигналов.

На вопрос о том, в какой степени скорость передачи информации может быть приближена к пропускной способности информационного канала, отвечает теорема Шеннона для дискретного канала без помех. Теорема формулируется следующим образом: если поток информации, вырабатываемый источником, достаточно близок к пропускной способности канала, т. е. если справедливо равенство

$$\overline{I}(X) = C - \delta, \qquad (5.13)$$

где δ — сколь угодно малая величина, то всегда можно найти такой способ кодирования, который обеспечит передачу всех сообщений, вырабатываемых источником, причем скорость передачи информации будет весьма близка к пропускной способности канала

$$\overline{I}(Z, Y) = C - \delta.$$

Обратное утверждение теоремы заключается в том, что невозможно обеспечить длительную передачу всех сообщений, если поток информации, вырабатываемый источником, превышает пропускную способность канала

$$\overline{I}(X) > C.$$

Таким образом, теорема Шеннона утверждает, что при выполнении условия (5.13) скорость передачи информации может быть в принципе сколь угодно приближена к пропускной способности канала. Это может быть обеспечено соответствующим кодированием сигналов. Однако рассмотренная теорема не отвечает на вопрос, каким образом нужно осуществлять кодирование.

5.3. Скорость передачи информации и пропускная способность дискретного канала с помехами

При наличии помех в канале передачи информации нарушается однозначное соответствие между входным и выходным алфавитами

канала. Одному входному сигналу могут соответствовать различные выходные сигналы (рис. 5.5, *б*). Вследствие случайного характера помех невозможно заранее точно установить, какой сигнал может быть принят на выходе канала при посылке определенного входного сигнала. Речь может идти только о вероятностях получения на выходе канала элементарного сигнала z_j при условии, что был отправлен элементарный сигнал y_i. Вероятностный характер связи между входным и выходным алфавитами канала передачи информации полностью определяется матрицей переходных вероятностей

$$\begin{vmatrix} p_{11} & p_{12} & p_{13} & \dots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \dots & p_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{n1} & p_{n2} & p_{n3} & \dots & p_{nn} \end{vmatrix},$$

где p_{ij} — условная вероятность перехода i-го символа входного алфавита в j-й символ выходного алфавита.

Очевидна справедливость следующего равенства:

$$\sum_{j=1}^{n} p_{ij} = 1.$$

Дискретный канал, по которому передаются только два элементарных сигнала, называется бинарным каналом. Матрица переходных вероятностей для такого канала имеет вид

$$\begin{vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{vmatrix}.$$

Если все вероятности правильной передачи сигналов одинаковы и одинаковы все вероятности искаженной передачи, то такой канал называется симметричным. Для симметричного бинарного канала матрица переходных вероятностей имеет вид

$$\begin{vmatrix} p & q \\ q & p \end{vmatrix},$$

где $p = p_{11} = p_{22}$ — вероятность правильной передачи; $q = p_{12} = p_{21}$ — вероятность искаженной передачи.

Так как в симметричном канале вероятности искажения всех символов сигнала одинаковы, то можно утверждать, что в таком канале помехи не зависят от передаваемых сигналов.

На рис. 5.6 приведен график переходных вероятностей двоичного симметричного канала.

Скорость передачи информации по дискретному каналу с помехами

$$\bar{I}(Z, Y) = \bar{U}_y [H(Y) - H(Y/Z)], \qquad (5.14)$$

где $H(Y/Z)$ — остаточная энтропия сигнала, обусловленная действием помех.

Выражение для скорости передачи информации может быть представлено также в виде

$$\bar{I}(Z, Y) = \bar{U}_y [H(Z) - H(Z/Y)], \qquad (5.15)$$

где $H(Z)$ — энтропия выходного сигнала; $H(Z/Y)$ — условная энтропия выходного сигнала при известной энтропии входного сигнала.

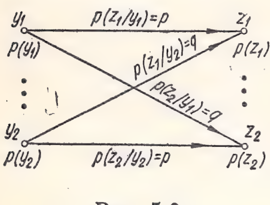


Рис. 5.6

Каналы, у которых на каждый передаваемый символ сигнала помехи воздействуют независимо от того, какие сигналы передавались ранее, называются каналами без памяти. В таких каналах помехи не вызывают дополнительных коррелятивных связей между символами. В настоящее время основные выводы теории информации получены применительно к каналам без памяти. Поэтому будем в дальнейшем рассматривать только каналы без памяти.

В случае независимости отдельных символов сигнала выражения (5.14) и (5.15) для канала без памяти примут вид

$$\bar{I}(Z, Y) = \bar{U}_y \Big[-\sum_{i=1}^{n} p(y_i) \log_2 p(y_i) + $$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} p(z_j) p(y_i/z_j) \log_2 p(y_i/z_j) \Big]; \qquad (5.16)$$

$$\bar{I}(Z, Y) = \bar{U}_y \Big[-\sum_{i=1}^{n} p(z_i) \log_2 p(z_i) + $$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} p(y_j) p(z_i/y_i) \log_2 p(z_i/y_j) \Big]. \qquad (5.17)$$

Пропускная способность канала с помехами может быть определена, исходя из выражения (5.6).

В качестве примера рассмотрим бинарный канал. Для такого канала алфавиты входного Y и выходного Z сигналов состоят из двух символов

$$Y = \{y_1, y_2\}; \quad Z = \{z_1, z_2\}.$$

Для кратности записи обозначим вероятности искажения сигналов

$$p(z_1/y_2) = q_1 \text{ и } p(z_2/y_1) = q_2.$$

Очевидно, что вероятности правильной передачи будут

$$p(z_1/y_1) = 1 - q_2; \quad p(z_2/y_2) = 1 - q_1.$$

При этом выражение (5.17) примет вид

$$\bar{I}(Z, Y) = \bar{U}_y \{-p(z_1)\log_2 p(z_1) - p(z_2)\log_2 p(z_2) + $$
$$+ p(y_1)[p(z_1/y_1)\log_2 p(z_1/y_1) + p(z_2/y_1)\log_2 p(z_2/y_1)] + $$
$$+ p(y_2)[p(z_1/y_2)\log_2 p(z_1/y_2) + p(z_2/y_2)\log_2 p(z_2/y_2)]\} = $$
$$= \bar{U}_y \{-p(z_1)\log_2 p(z_1) - p(z_2)\log_2 p(z_2) + p(y_1)[(1-q_2)\log_2(1-q_2) + $$
$$+ q_2\log_2 q_2] + p(y_2)[q_1\log_2 q_1 + (1-q_1)\log_2(1-q_1)]\}.$$

Максимизируя правую часть выражения (5.17), мы сможем определить пропускную способность канала. Очевидно, что этого можно достичь за счет оптимизации значений априорных вероятностей $p(y_1)$ и $p(y_2)$ передачи сигналов y_1 и y_2, так как никакие другие параметры канала менять мы не можем.

Рассмотрим частные случаи.

1. Вероятности искажений сигналов $q_1 = q_2 = q$. Этот случай соответствует симметричному каналу. Для симметричного канала условная энтропия

$$H(Z/Y) = -[p(y_1) + p(y_2)][q\log_2 q + (1-q)\log_2(1-q)] = $$
$$q\log_2 q + (1-q)\log_2(1-q), \qquad (5.18)$$

так как $p(y_1) + p(y_2) = 1$.

Из выражения (5.18) видно, что условная энтропия не зависит от априорных вероятностей $p(y_1)$ и $p(y_2)$. Следовательно, максимальная скорость передачи информации получается в этом случае при таком распределении вероятностей $p(y_1)$ и $p(y_2)$, при котором энтропия $H(Z)$ оказывается максимальной. Это будет иметь место при равенстве априорных вероятностей $p(z_1) = p(z_2)$. При этом максимальное значение энтропии $H(Z)_{\text{макс}} = 1$ дв. ед. и пропускная способность канала определится выражением

$$C = \bar{U}_y[\log_2 2 + q\log_2 q + (1-q)\log_2(1-q)]. \qquad (5.19)$$

Используя известное правило теории вероятностей

$$p(z_j) = \sum_{i=1}^{n} p(y_i)\,p(z_i/y_i), \qquad (5.20)$$

можно показать, что равенство априорных вероятностей выходных сигналов z_1 и z_2 для симметричного канала будет иметь место при равенстве априорных вероятностей входных сигналов, т. е. при $p(y_1) = p(y_2)$.

Таким образом, в симметричном бинарном канале с помехами максимальная скорость передачи информации получается при таком же условии, как и в канале без помех. Однако из сравнения (5.19) и (5.12) видно, что наличие помех в канале приводит к уменьшению пропускной способности канала.

2. Вероятности искажений сигналов $q_1 = 0$, $q_2 \neq 0$, т. е. имеют место искажения лишь при передаче сигналов y_2.

Из (5.17) с учетом (5.20) для этого случая получаем

$$\bar{I}(Z,\ Y) = \bar{U}_y \{-p(y_1)\log_2 p(y_1)\, q_2 \log_2 q_2 - $$
$$-[p(y_2) + p(y_1)\, q_2]\log_2 [p(y_2) + p(y_1)\, q_2]\}.$$

Максимальное значение скорости передачи информации будет при условии [10]

$$p(y_1)_{\text{опт}} = \frac{1}{1 + q_2^{\frac{q_2}{1-q_2}} - q_2}. \qquad (5.21)$$

Пропускная способность такого канала определяется выражением

$$C = \bar{U}_y \{-p(y_1)_{\text{опт}} + \log_2 p(y_1)_{\text{опт}} + p(y_1)_{\text{опт}}\, q_2 \log_2 q_2 -$$
$$-[1 - p(y_1)_{\text{опт}} + p(y_1)_{\text{опт}}\, q_2]\log_2 [1 - p(y_1)_{\text{опт}} + p(y_1)_{\text{опт}}\, q_2]\}. \qquad (5.22)$$

На рис. 5.7 приведены графики $\frac{C}{\bar{U}_y} = C\bar{\tau}_y$, $p(y_1)_{\text{опт}}$ и $p(y_2)_{\text{опт}} = 1 - p(y_1)_{\text{опт}}$ как функции от вероятности q_2 [10]. Как видно из графиков, с увеличением q_2 от 0 до 1 пропускная способность падает от $\bar{U}_y = \frac{1}{\tau_y}$ до нуля. Для дискретного канала с помехами Шенноном доказана следующая теорема: если поток информации, вырабатываемый источником, достаточно близок к пропускной способности канала, т. е. если справедливо равенство

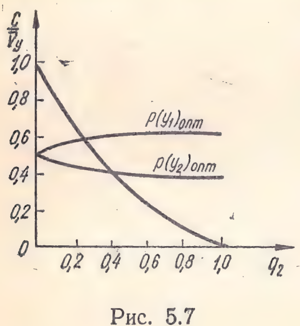


Рис. 5.7

$$\bar{I}(X) = C - \delta,$$

где δ — сколь угодно малая величина, то всегда можно найти такой способ кодирования, который обеспечит передачу всех сообщений, вырабатываемых источником, а вероятность ошибочного опознания любого переданного сообщения будет сколь угодно малой, т. е. $p_{\text{н. о}} < \eta$, где $p_{\text{н. о}}$ — вероятность неправильного опознания переданного сообщения; η — сколь угодно малая величина.

Обратное утверждение теоремы состоит в том, что если поток информации источника превышает пропускную способность канала, то не существует способа кодирования, обеспечивающего передачу любого сообщения с малой вероятностью ошибки.

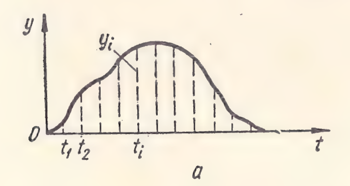
Таким образом, рассмотренная теорема определяет соотношение между скоростью создания сообщений источником, пропускной способностью канала при наличии помех и достоверностью передачи.

Если для канала без помех представляет интерес эффективность передачи, то для канала с помехами — как эффективность, так и достоверность передачи.

Данная теорема, как и теорема для канала без помех, не отвечает на вопрос, каким образом нужно осуществлять кодирование, чтобы приблизить скорость передачи информации к пропускной способности канала. Но для приближения скорости передачи к пределу общим методом как для канала с помехами, так и для канала без помех является кодирование длинных сообщений [10, 28].

5.4. Скорость передачи информации и пропускная способность непрерывного канала с помехами

Под непрерывным каналом передачи информации принято понимать совокупность средств, предназначенных для передачи непрерывных сигналов [2]. В отличие от дискретных каналов в непре-



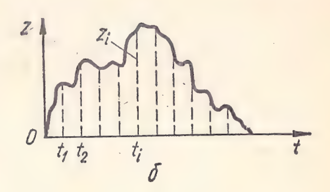


Рис. 5.8

рывных каналах вместо кодирующих и декодирующих устройств может использоваться более широкий класс различных преобразователей. Для передачи информации по каналу может применяться модуляция одного или нескольких параметров сигнала. Независимо от конкретного характера преобразования сигналов входные и выходные сигналы непрерывного канала задаются в виде ансамблей непрерывных функций с соответствующими плотностями распределения вероятностей.

Пусть на вход канала поступает непрерывный сигнал $Y(t)$ длительностью T. Вследствие воздействия помех $\xi(t)$ выходной сигнал $Z(t)$ будет отличаться от входного.

Количество информации, содержащееся в случайном сигнале $Z(t)$ о случайном сигнале $Y(t)$, определяется известной зависимостью

$$I_T(Z,\ Y) = H_T(Z) - H_T(Z/Y). \tag{5.23}$$

В соответствии с теоремой Котельникова, непрерывные сигналы $Y(t)$ и $Z(t)$ могут быть представлены совокупностями отсчетов y_i и z_i в дискретные моменты времени (рис. 5.8), представляющих собой случайные величины.

Распределение совокупности случайных величин описывается многомерными плотностями распределения вероятности

$$w(y_1,\ y_2,\ \ldots,\ y_m) \text{ и } w(z_1,\ z_2,\ \ldots,\ z_m).$$

Тогда дифференциальная энтропия сигнала на выходе канала

$$h_T(Z) = -\int\limits_{-\infty}^{\infty} \ldots m \ldots \int\limits_{-\infty}^{\infty} w(z_1, z_2, \ldots, z_m) \log_2 w(z_1, z_2, \ldots, z_m)\, dz_1, dz_2 \ldots dz_m.$$

Так как в соответствии с критерием Н. А. Железнова (3.9) при квантовании случайных сигналов по времени интервал квантования необходимо устанавливать равным интервалу корреляции функции τ_0, то случайные величины $z_1, z_2, \ldots, z_i, \ldots, z_m$ можно полагать независимыми и

$$w(z_1, z_2, \ldots, z_m) = w(z_1)\, w(z_2) \ldots w(z_m).$$

Исходя из того, что энтропия совокупности случайных независимых величин равна сумме энтропий случайных величин, получим следующее выражение для дифференциальной энтропии сигнала:

$$h_T(Z) = \sum_{i=1}^{m} h(z_i),$$

где $h(z_i) = -\int\limits_{-\infty}^{\infty} w(z_i) \log_2 w(z_i)\, dz_i$ — дифференциальная энтропия i-го отсчета сигнала Z; $m = \frac{T}{\Delta T}$ — общее количество отсчетов сигнала Z длительностью T; ΔT — интервал временного квантования.

Ограничиваясь рассмотрением стационарных процессов, получим

$$w(z_1) = w(z_2) = \cdots = w(z_m);$$
$$h(z_1) = h(z_2) = \cdots = h(z_m) = h(Z).$$

Тогда

$$h_T(Z) = mh(Z),$$

где $h(Z)$ — дифференциальная энтропия одного отсчета.

Аналогично можно показать, что условная дифференциальная энтропия

$$h_T(Z/Y) = mh(Z/Y),$$

где $h(Z/Y)$ — условная дифференциальная энтропия одного отсчета.

Таким образом, выражение для количества информации принимает вид

$$I_T(Z, Y) = m[h(Z) - h(Z/Y)].$$

Скорость передачи информации

$$\overline{I}_T(Z, Y) = \frac{m}{T}[h(Z) - n(Z/Y)] = F_0[h(Z) - h(Z/Y)], \quad (5.24)$$

где $F_0 = \frac{m}{T} = \frac{1}{\Delta T}$ — частота временного квантования (отсчета).

Пропускная способность канала

$$C = \text{макс}\,[\bar{I}_T(Z,\ Y)] = F_0\,\text{макс}\,[h(Z) - h(Z/Y)]. \qquad (5.25)$$

Максимальное значение правой части выражения (5.25) можно достичь, варьируя $w(Y)$, так как остальные параметры относятся к каналу связи, которые мы менять не можем.

Рассмотрим некоторые частные случаи [27].

1. Сигнал ограниченной мощности передается по каналу, в котором действует аддитивная помеха ограниченной мощности типа белого гауссового шума.

При аддитивной помехе сигнал $Z(t)$ на выходе канала будет равен

$$Z(t) = Y(t) + \xi(t),$$

где $\xi(t)$ — помеха, действующая в канале передачи информации.

Средние мощности сигнала и помехи соответственно равны P_y и $P_\xi = \sigma_\xi^2$. Полоса пропускания канала ограничена пределами 0 и $F_к$.

Ширина спектра сигнала и помехи ограничиваются полосой пропускания канала.

Частота временного квантования ограниченного по спектру сигнала в соответствии с теоремой Котельникова

$$F_0 = 2F_к.$$

Тогда выражение (5.25) для пропускной способности канала примет вид

$$C = 2F_к\,\text{макс}\,[h(Z) - h(Z/Y)]. \qquad (5.26)$$

При взаимно независимых сигнале Y и помехе ξ вероятность того, что при передаче сигнала Y выходной сигнал будет равен $Z = Y + \xi$, должна определяться вероятностью того, что помеха примет данное значение $\xi = Z - Y$, т. е.

$$w(Z/Y)\,dY = w(\xi)\,d\xi,$$

а

$$w(Z/Y) = w(Y + \xi/Y) = w(Y/Y + \xi/Y) = \\ = w(Y/Y)\cdot w(\xi/Y) = w(\xi). \qquad (5.27)$$

Учитывая (5.27), выражение для условной дифференциальной энтропии $h(Z/Y)$ можно преобразовать следующим образом:

$$h(Z/Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(Y)\,w(Z/Y)\log_2 w(Z/Y)\,dY\,dZ =$$

$$= -\int_{-\infty}^{\infty} w(Y)\left[\int_{-\infty}^{\infty} w(\xi)\log_2 w(\xi)\,d\xi\right]dY =$$

$$= h(\xi)\int_{-\infty}^{\infty} w(Y)\,dY = h(\xi), \qquad (5.28)$$

где $h(\xi)$ — дифференциальная энтропия помехи.

Таким образом, в случае аддитивной помехи условная дифференциальная энтропия $h(Z/Y)$ полностью определяется свойствами помехи.

Ранее было установлено (см. пример 4.3), что дифференциальная энтропия сигнала, распределенного по нормальному закону,

$$h(\xi) = \log_2 (\sqrt{2\pi e}\sigma_\xi). \qquad (5.29)$$

Подставив (5.29) в (5.26), получим следующее выражение для пропускной способности канала

$$C = 2F_к \text{ макс } \{h(Z) - \log_2 \sqrt{2\pi e}\sigma_\xi\}. \qquad (5.30)$$

Так как σ_ξ задано, то максимальное значение выражения (5.30) будет обеспечено при условии максимизации дифференциальной энтропии выходного сигнала $h(Z)$. Так как средние мощности входного сигнала $Y(t)$ и помехи $\xi(t)$ ограничены, то средняя мощность выходного сигнала $Z(t)$ также ограничена. При таком условии величина дифференциальной энтропии $h(Z)$ будет максимальной в случае, если $Z(t)$ характеризуется нормальным законом распределения. Если же суммарный сигнал $Z(t)$ и одна из его составляющих $\xi(t)$ распределены по нормальному закону, то и вторая составляющая, т. е. входной сигнал $Y(t)$, также должна иметь нормальный закон распределения.

Таким образом, дифференциальная энтропия выходного сигнала

$$h(Z) = \log_2 \sqrt{2\pi e}\sigma_Z = \log_2 \sqrt{(\sigma_y^2 + \sigma_\xi^2) 2\pi e}. \qquad (5.31)$$

Подставляя (5.31) в (5.30), окончательно получим

$$C = 2F_к [\log_2 \sqrt{(\sigma_y^2 + \sigma_\xi^2) 2\pi e} - \log_2 \sqrt{\sigma_\xi^2 2\pi e}] =$$
$$= F_к \log_2 \frac{\sigma_y^2 + \sigma_\xi^2}{\sigma_\xi^2} = F_к \log_2 \left(1 + \frac{P_y}{P_\xi}\right), \qquad (5.32)$$

где $P_y = \sigma_y^2$ — средняя мощность полезного сигнала; $P_\xi = \sigma_\xi^2$ — средняя мощность помехи.

Итак, скорость передачи информации сигналами с ограниченной средней мощностью через канал, в котором действует белый гауссов шум, оказывается максимальной при полном подобии между сигналом и помехой. Следовательно, максимальная скорость передачи информации будет обеспечена, если в качестве физического переносчика информации применять стационарный случайный процесс в виде белого гауссова шума.

Как видно (5.32), пропускную способность канала можно регулировать путем изменения $F_к$ или P_y. Причем практически зависимость пропускной способности канала от $F_к$ при постоянной мощности сигнала нелинейная. Это обусловлено тем, что мощность помехи P_ξ также зависит от ширины частотного спектра. Действи-

тельно, энергетический спектр белого шума равномерен, поэтому мощность такой помехи можно представить в виде

$$P_\xi = P_0 F_к, \qquad (5.33)$$

где P_0 — мощность помехи, приходящаяся на полосу в 1 Гц (спектральная плотность мощности помехи).

Подставив (5.33) в (5.32), получим выражение, определяющее действительный характер зависимости пропускной способности канала от ширины его полосы пропускания

$$C = F_к \log_2\left(1 + \frac{P_y}{P_0 F_к}\right). \qquad (5.34)$$

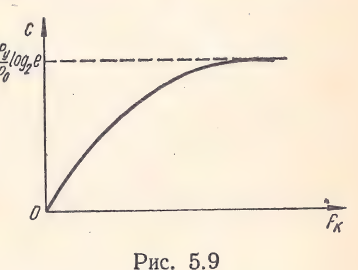


Рис. 5.9

Характер зависимости $C = f(F_к)$ представлен графически на рис. 5.9.

Определим предел, к которому стремится пропускная способность канала при неограниченном увеличении его полосы пропускания

$$\lim_{F_к \to \infty} C = \lim_{F_к \to \infty} \frac{\log_2\left(1 + \frac{P_y}{P_0 F_к}\right)}{\frac{1}{F_к}}.$$

Введя обозначение $\alpha = \frac{1}{F_к}$, получим

$$\lim_{F_к \to \infty} C = \lim_{\alpha \to 0} \frac{\log_2\left(1 + \frac{P_y}{P_0}\alpha\right)}{\alpha}.$$

Раскрывая неопределенность, получим предельное значение пропускной способности канала

$$\lim_{F_к \to \infty} C = \frac{P_y}{P_0}\log_2 e. \qquad (5.35)$$

Из формулы (5.35) видно, что максимальное значение, к которому стремится пропускная способность канала с ростом его ширины полосы пропускания, пропорционально отношению средней мощности сигнала к спектральной плотности мощности помехи.

Из произведенного анализа можно также заключить, что нет смысла сильно увеличивать полосу пропускания канала, так как по мере расширения полосы пропускания рост пропускной способности канала замедляется и в пределе при $F_к \to \infty$ пропускная способность приближается к постоянной величине. Имеет смысл увеличивать полосу пропускания примерно до значения, равного отношению $\frac{P_y}{P_0}$. Предельно возможное значение пропускной способности может быть увеличено за счет увеличения отношения $\frac{P_y}{P_0}$.

2. Сигнал ограниченной мощности передается по каналу, в котором действует аддитивная помеха в виде произвольного шума. Как уже установлено, при определенном среднеквадратическом значении (при определенной мощности) помехи наибольшей энтропией обладает помеха с нормальным законом распределения вероятностей. При любом другом законе распределения вероятностей помехи ее энтропия (при том же среднеквадратическом значении) оказывается меньшей. Зависимость энтропии от вида закона распределения побудила Шеннона характеризовать действие помехи не ее действительной мощностью, а так называемой энтропийной мощностью помехи. Под энтропийной мощностью Шеннон понимал мощность эквивалентного белого шума, который, имея те же длительность и ширину спектра, обладает такой же, как данная помеха, энтропией.

Таким образом, если произвольная помеха $\xi(t)$ характеризуется энтропией на один отсчет $h(\xi)$, то тогда мощность эквивалентного белого шума, т. е. энтропийная мощность $\sigma^2_{\xi э}$, может быть определена из условия

$$h(\xi) = \log_2 \sqrt{2\pi e \sigma^2_{\xi э}},$$

откуда

$$\sigma^2_{\xi э} = \frac{2^{h(\xi)}}{2\pi e} = \frac{2^{h(\xi)-1}}{\pi e}. \tag{5.36}$$

Величина отношения средней мощности помехи σ^2_ξ и ее энтропийной мощности $\sigma^2_{\xi э}$ определяется законом распределения помехи. Обозначив это отношение через $k_э$, получим

$$\sigma^2_\xi = k_э \sigma^2_{\xi э}. \tag{5.37}$$

В частности, для закона равной вероятности $k_э = 1{,}3$.

Используя понятие энтропийной мощности, можно выражение для пропускной способности канала, в котором действует произвольная помеха, представить в следующем виде:

$$C = 2F_к \left[\log_2 \sqrt{(\sigma^2_y + \sigma^2_\xi)\, 2\pi e} - \log_2 \sqrt{\sigma^2_{\xi э} 2\pi e}\right] =$$
$$= F_к \log_2 \left(\frac{\sigma^2_y + \sigma^2_\xi}{\sigma^2_{\xi э}}\right) = F_к \log_2 \left(k_э \frac{\sigma^2_y + \sigma^2_\xi}{\sigma^2_\xi}\right). \tag{5.38}$$

Для всех законов распределения (кроме нормального) коэффициент $k_э > 1$, поэтому пропускная способность канала, в котором действует произвольная помеха, будет всегда больше пропускной способности канала, в котором действует белый шум с такой же средней мощностью, что и произвольная помеха.

Для непрерывного канала с помехами Шенноном сформулирована следующая теорема:

Если энтропия $\overline{H}_\varepsilon(X)$ источника непрерывных сообщений, определяющая количество информации, вырабатываемое в единицу вре-

мени при заданной оценке g верности воспроизведения, сколь угодно близка к пропускной способности канала, т. е. справедливо соотношение

$$\overline{H}_{\varepsilon}(X) = C - \alpha,$$

где α — как угодно мало, то существует метод передачи, при котором все сообщения, вырабатываемые источником, могут быть переданы, а верность воспроизведения при этом как угодно близка к g.

Обратное утверждение этой теоремы говорит о том, что такая передача невозможна, если

$$\overline{H}_{\varepsilon}(X) > C.$$

Теорема позволяет находить предельно достижимую эффективность непрерывных каналов.

В заключение рассмотрим связь между пропускной способностью и емкостью канала передачи информации (см. 5.1).

Количество информации, которое может быть передано по каналу за время его работы $T_к$ при воздействии помех типа белого шума,

$$I_T(Z, Y) = T_к F_к \log_2\left(1 + \frac{P_y}{P_\xi}\right). \qquad (5.39)$$

Практически в большинстве случаев мощность полезного сигнала значительно превышает мощность помех. В этих случаях выражение (5.39) можно с достаточным приближением представить в виде

$$I_T(Z, Y) = T_к F_к \log_2 \frac{P_y}{P_\xi}. \qquad (5.40)$$

В формулах (5.39) и (5.40) члены $\log_2\left(1 + \frac{P_y}{P_\xi}\right)$ и $\log_2 \frac{P_y}{P_\xi}$ выражают (с точностью до постоянного множителя) максимально возможное количество информации на один отсчет. Тогда, сравнивая (5.39) и (5.40) с (5.2), можно заключить, что емкость канала определяет максимально возможное количество информации, которое может быть передано по каналу за время его работы. А так как пропускная способность выражает максимально возможное количество информации, которое может быть передано по каналу за единицу времени, то связь между емкостью и пропускной способностью канала определяется зависимостью

$$V_к = T_к C. \qquad (5.41)$$

5.5. Примеры

Пример 5.1. В информационном канале без помех для передачи сообщений используется алфавит с четырьмя различными символами. Длительности всех символов одинаковы и равны $\tau = 1$ мс. Определить пропускную способность канала передачи информации.

Решение. Для расчета пропускной способности дискретного канала без помех воспользуемся формулой

$$C = \overline{U}_y \text{макс} \{H(Y)\} = \frac{\log_2 N}{T_y},$$

где N — общее количество сообщений; T_y — средняя длительность сигнала.

Так как код каждого сообщения содержит четыре символа, то длительность всех сигналов будет постоянна и равна

$$T_y = 4\tau = 4 \text{ мс}.$$

Так как для передачи сообщений используется алфавит с четырьмя символами, то

$$N = 4^4.$$

Следовательно,

$$C = 250 \log_2 4^4 = 250 \log_2 2^8 = 2 \cdot 10^3 \frac{\text{дв. ед.}}{\text{с}}.$$

Пример 5.2. Источник вырабатывает символы с вероятностями

$$p_1 = 0{,}2; \quad p_2 = 0{,}7 \quad \text{и} \quad p_3 = 0{,}1;$$

корреляционные связи между сообщениями отсутствуют. Передача информации осуществляется двоичным кодом, длительность символов которого равна $\tau = 1$ мс. Определить скорость передачи информации по каналу без помех при использовании равномерного кода.

Решение. Скорость передачи информации определяем по формуле (5.11).

Средняя энтропия сообщений на один символ

$$H(Y) = -\sum_{i=1}^{3} p_i \log_2 p_i = -0{,}2 \log_2 0{,}2 - 0{,}7 \log_2 0{,}7 -$$
$$- 0{,}1 \log_2 0{,}1 = 1{,}16 \frac{\text{дв. ед.}}{\text{симв.}}.$$

Для передачи трех сообщений двоичным кодом необходимо два разряда. Следовательно, длина кодовых комбинаций равна 2τ и скорость передачи сигналов $U_y = \frac{1}{2\tau} = 500 \frac{\text{сообщ.}}{\text{с}}$.

Таким образом, скорость передачи информации

$$\overline{I}(Y) = U_y H(Y) = 500 \cdot 1{,}16 = 580 \frac{\text{дв. ед.}}{\text{с}}.$$

Пример 5.3. Источник, вырабатывающий четыре символа с априорными вероятностями $p_1 = 0{,}4$; $p_2 = 0{,}3$; $p_3 = 0{,}2$ и $p_4 = 0{,}1$, подключен к каналу передачи информации, обладающему пропускной способностью $C = 1000 \frac{\text{дв. ед.}}{\text{с}}$. Передача информации осуществляется равномерным двоичным кодом.

С какой скоростью будет осуществляться передача информации?

Решение. Скорость передачи информации определяем по формуле (5.11).

Средняя энтропия сообщений на один символ источника

$$H(Y) = -0{,}4\log_2 0{,}4 - 0{,}3\log_2 0{,}3 - 0{,}2\log_2 0{,}2 - 0{,}1\log_2 0{,}1 = \\ = 1{,}848\ \frac{\text{дв. ед.}}{\text{симв.}}.$$

Скорость передачи символов U_y может быть определена из выражения для пропускной способности канала. Действительно, максимальная энтропия будет иметь место при равной вероятности и статистической независимости сообщений

$$\text{макс}\{H(Y)\} = \log_2 4 = 2\,\frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Тогда из (5.12) получим

$$\overline{U}_y = \frac{C}{\text{макс}\{H(Y)\}} = \frac{1000}{2} = 500\ \frac{\text{симв.}}{\text{с}}.$$

Следовательно, скорость передачи информации

$$\overline{I}(\overline{Y}) = \overline{U}_y H(Y) = 500 \cdot 1{,}848 = 924\,\frac{\text{дв. ед.}}{\text{с}}$$

Пример 5.4. Сколько в среднем можно передать букв русского текста в секунду по каналу с пропускной способностью $C = 1000\,\frac{\text{дв. ед.}}{\text{с}}$.

Решение. В 4.7 установлено, что средняя энтропия русского языка на одну букву с учетом всех коррелятивных связей

$$H(Y) = 2\ \frac{\text{дв. ед.}}{\text{букву}}.$$

Следовательно, средняя скорость передачи русского текста по каналу

$$\overline{U}_y = \frac{1000}{2} = 500\ \frac{\text{букв}}{\text{с}}.$$

Пример 5.5. При передаче непрерывных сигналов по каналу связи осуществляется кодоимпульсная модуляция. Квантование по времени осуществляется с частотой $F_к = 100$ Гц. При квантовании по уровню шкала уровней сигнала делится на 64 дискретных уровня.

Определить пропускную способность канала.

Решение. Количество дискретных уровней определяет количество передаваемых сообщений. Следовательно, максимально возможная энтропия сообщений

$$\text{макс}\{H(Y)\} = \log_2 64 = 6\ \frac{\text{дв. ед.}}{\text{сообщ.}},$$

а пропускная способность канала

$$C = U_y\,\text{макс}\{H(Y)\} = F_к\,\text{макс}\{H(Y)\} = 100 \cdot 6 = 600\ \frac{\text{дв. ед.}}{\text{с}}.$$

Пример 5.6 (из [27]). По каналу связи с помехами передаются три сообщения, которым соответствуют сигналы y_1, y_2 и y_3. При этом канал таков, что сигнал y_1 всегда передается безошибочно, а сигналы y_2 и y_3 передаются или правильно с вероятностями p, или с искажениями, когда один сигнал переходит в другой с вероятностью q. Априорные вероятности посылки сигналов $p(y_1) = p$; $p(y_2) = p(y_3) = Q$. Схема передачи сигналов показана на рис. 5.10.

Определить пропускную способность канала, если скорость передачи сигналов равна U_y.

Рис. 5.10

Решение. Пропускная способность дискретного канала с помехами может быть определена по формуле

$$C = \overline{U}_y \text{макс}\,\{H(Y) - H(Y/X)\}.$$

Начальная энтропия сигналов

$$H(Y) = -\sum_{i=1}^{3} p(y_i)\log_2 p(y_i) = -[p\log_2 p + 2Q\log_2 Q].$$

Остаточная энтропия определяется выражением

$$H(Y/Z) = -\sum_{i=1}^{3}\sum_{j=1}^{3} p(z_i)\,p(y_j/z_i)\log_2 p(y_j/z_i) =$$
$$= -p(z_1)\,p(y_1/z_1)\log_2 p(y_1/z_1) - p(z_2)\,[p(y_2/z_2)\log_2 p(y_2/z_2) +$$
$$+ p(y_3/z_2)\log_2 p(y_3/z_2)] - p(z_3)\,[p(y_2/z_3)\log_2 p(y_2/z_3) +$$
$$+ p(y_3/z_3)\log_2 p(y_3/z_3)].$$

Так как по условию задачи $p(y_1/z_1) = 1$, то первое слагаемое в нашем выражении будет равно нулю.

Кроме того, по условию задачи $p(y_2/z_2) = p(y_3/z_3) = p$; $p(y_2/z_3) = p(y_3/z_2) = q$.

Вероятности $p(z_2)$ и $p(z_3)$ находим из формулы полной вероятности

$$p(z_2) = p(y_2)\,p(z_2/y_2) + p(y_3)\,p(z_2/y_3) = Q\,[p(z_2/y_2) + p(z_2/y_3)] = Q;$$
$$p(z_3) = p(y_2)\,p(z_3/y_2) + p(y_3)\,p(z_3/y_3) = Q\,[p(z_3/y_2) + p(z_3/y_3)] = Q.$$

Следовательно, остаточная энтропия

$$H(Y/Z) = -2Q\,(p\log_2 p + q\log_2 q).$$

Таким образом, пропускная способность канала определяется выражением

$$C = \overline{U}_y \text{макс}\,\{-p\log_2 p - 2Q\log_2 Q + 2Q\,(p\log_2 p + q\log_2 q)\}.$$

Искомая величина C является функцией априорных вероятностей p и Q. Следовательно, нахождение максимума правой части полученного выражения сводится к соответствующему выбору p и Q при условии, что $p + 2Q = 1$. Как известно, такие задачи на мак-

симум решаются с помощью неопределенных коэффициентов Лагранжа. В соответствии с этим методом составим новую функцию вида

$$F = p \log_2 p - 2Q \log_2 Q + 2Q (p \log_2 p + q \log_2 q) + \lambda (p + 2Q),$$

где λ — коэффициент Лагранжа.

Для удобства математических выкладок перейдем к натуральным логарифмам

$$F = -p \ln p - 2Q \ln Q + 2Q (p \ln p + q \ln q) + \lambda (p + 2Q).$$

Ищем максимум данной функции, для чего найдем предварительно ее частные производные по p и Q и приравняем их к нулю

$$\frac{\partial F}{\partial p} = -1 - \ln p + \lambda = 0;$$

$$\frac{\partial F}{\partial Q} = -2 - 2 \ln Q + 2 (p \ln p + q \ln q) + 2\lambda = 0.$$

Из полученных уравнений найдем

$$p = e^{\lambda - 1};$$

$$Q = e^{\lambda - 1} \cdot e^{(p \ln p + q \ln q)}.$$

Из условия $p + 2Q = 1$ найдем

$$e^{\lambda - 1} (1 + 2e^{(p \ln p + q \ln q)}) = 1,$$

откуда

$$e^{\lambda - 1} = \frac{1}{1 + 2e^{(p \ln p + q \ln q)}} = \frac{\delta}{\delta + 2},$$

где

$$\delta = e^{-(p \ln p + q \ln q)}.$$

Таким образом, оптимальные значения априорных вероятностей равны

$$p = \frac{\delta}{\delta + 2};$$

$$Q = \frac{1}{\delta + 2}.$$

Следовательно, искомое значение пропускной способности канала

$$C = \overline{U}_у \left[-\frac{\delta}{\delta + 2} \ln \frac{\delta}{\delta + 2} - \frac{2}{\delta + 2} \ln \frac{1}{\delta + 2} - \frac{2}{\delta + 2} \ln \delta \right] =$$

$$= \overline{U}_у \left[-\ln \delta \left(\frac{\delta}{\delta + 2} + \frac{2}{\delta + 2} \right) + \ln (\delta + 2) \left(\frac{\delta}{\delta + 2} + \frac{2}{\delta + 2} \right) \right] =$$

$$= \overline{U}_у [\ln (\delta + 2) - \ln \delta] = \overline{U}_у \ln \frac{\delta + 2}{\delta}.$$

Пример 5.7. Определить пропускную способность непрерывного канала с помехами типа белого гауссова шума. Полоса пропускания канала $F_к = 1000$ Гц. Дисперсия полезного сигнала $\sigma_у^2 = P_у = 4000$ В^2, спектральная плотность мощности помехи $P_0 = 0{,}002$ В^2/с.

Решение. Для определения пропускной способности непрерывного канала с помехами воспользуемся формулой (5.34)

$$C = F_к \log_2\left(1 + \frac{P_у}{P_0 F_к}\right) = 1000 \log_2\left(1 + \frac{4 \cdot 10^3}{0{,}002 \cdot 10^3}\right) = 10\,970\ \frac{\text{дв. ед.}}{\text{с}}.$$

Пример 5.8. Определить скорость передачи информации сигналом, имеющим среднюю мощность $P_у = 4000$ В^2, длительность $T = 20$ м и ограниченным полосой частот $F_c = 1000$ Гц. При передаче сигнал подвергается амплитудно-импульсной модуляции (рис. 5.11). Величина сигнала распределена равновероятно по всей шкале уровней. Дисперсия погрешности квантования по уровню не должна превышать $\sigma_к^2 = 17$ В^2.

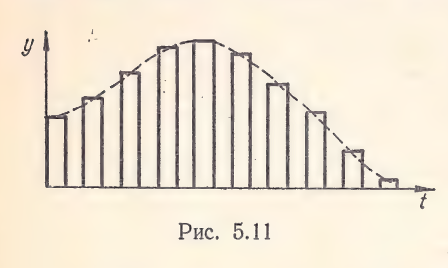


Рис. 5.11

Решение. Скорость передачи информации определяем по формуле

$$\overline{I}(Y) = \overline{U}_у \log_2 N = F_c \log_2 N^2,$$

где N — количество уровней квантования сигнала; $\overline{U}_у = 2F_c$ — частота квантования сигнала во времени.

Количество уровней квантования может быть выражено через среднюю мощность сигнала $P_у$. Действительно, средняя мощность сигнала при равной вероятности уровней квантования

$$P_у = \frac{1}{N}\sum_{i=1}^{N} i^2 \Delta_у^2 = \frac{\Delta_у^2}{N}\sum_{i=1}^{N} i^2,$$

где $\Delta_у$ — величина шага квантования сигнала по уровню.

Как известно, сумма ряда

$$\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}.$$

При $N \gg 1$

$$\sum_{i=1}^{N} i^2 \cong \frac{N^3}{6}.$$

Тогда

$$P_у \cong \frac{N^2 \Delta_у^2}{3}.$$

Шаг квантования по уровню связан со среднеквадратической погрешностью квантования зависимостью (см. 3.10)

$$\sigma_к = \frac{\Delta_y}{2\sqrt{3}}.$$

Следовательно,

$$P_y = 4N^2\sigma_k^2,$$

откуда

$$N^2 = \frac{P_y}{4\sigma_k^2} = 60.$$

Таким образом, скорость передачи информации

$$\overline{I}(Y) = 1000 \log_2 60 = 5910 \ \frac{\text{дв. ед.}}{\text{с}}.$$

Пример 5.9 (из [33]). Определить необходимую пропускную способность и полосу пропускания канала, предназначенного для передачи телевизионного изображения. Можно полагать телевизионное изображение состоящим из 300 000 мелких элементов изображения. Каждый из этих элементов может принимать 10 различных градаций яркости. Все градации можно считать равновероятными. За одну секунду передается 30 кадров изображения. Кроме того, известно, что для удовлетворительного воспроизведения изображения необходимое отношение сигнал/шум равно 1000.

Решение. Поскольку каждый элемент изображения может принимать 10 уровней с равной вероятностью, то один элемент изображения содержит количество информации

$$i = \log_2 10 = 3{,}32 \ \frac{\text{дв. ед.}}{\text{элемент}}.$$

Один кадр содержит количество информации, равное $300\,000i$. Поскольку за одну секунду должно быть передано 30 кадров, то пропускная способность канала

$$C = 30 \cdot 300\,000i = 29{,}9 \cdot 10^6 \ \frac{\text{дв. ед.}}{\text{с}}.$$

С другой стороны, при известном отношении сигнал/шум пропускная способность канала может быть определена по формуле (5.32), откуда полоса пропускания канала

$$F_\text{к} \cong \frac{C}{\log_2 \frac{P_\text{y}}{P_\xi}} = \frac{29{,}9 \cdot 10^6}{9{,}95} \cong 3 \text{ мГц}.$$

Пример 5.10. Определить пропускную способность телефонного канала, обеспечивающего передачу разговора со скоростью $100 \ \frac{\text{слов}}{\text{мин}}$. Среднюю длину слова считать состоящей из 5 букв.

Решение. Считая, что на каждую букву в среднем приходятся две двоичные единицы информации (см. 4.7), получим следующее значение пропускной способности канала:

$$C = \frac{100 \cdot 5 \cdot 2}{60} = 16{,}6 \ \frac{\text{дв. ед.}}{\text{с}}.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какими обобщенными характеристиками определяется сигнал и канал передачи информации?
2. Что понимается под объемом сигнала и емкостью канала передачи информации?
3. Сформулируйте необходимые и достаточные условия согласования сигнала с каналом передачи информации?
4. Чем вызвана необходимость согласования сигнала с каналом передачи информации?
5. Объясните, как можно передать широкополосный сигнал через узкополосный канал без искажений.
6. Что понимается под скоростью передачи информации и пропускной способностью канала?
7. Какими факторами определяется скорость передачи информации и пропускная способность канала?
8. Что понимается под дискретным каналом и непрерывным каналом?
9. В чем состоит сущность теоремы Шеннона для дискретного канала без помех?
10. Что характеризует матрица переходных вероятностей?
11. Какими свойствами обладает симметричный бинарный канал?
12. Что понимается под каналом без памяти?
13. Каким образом вычисляется скорость передачи информации по дискретному каналу с помехами и без помех?
14. Каким образом определяется пропускная способность дискретного канала с помехами и без помех?
15. В чем сущность теоремы Шеннона для дискретного канала с помехами?
16. Каким образом определяется скорость передачи информации по непрерывному каналу с помехами?
17. Каким образом определяется пропускная способность непрерывного канала с помехами при передаче сигналов с ограниченной мощностью?
18. Объясните характер зависимости пропускной способности непрерывного канала с помехами от полосы пропускания канала?
19. Что понимается под энтропийной мощностью помехи?
20. В чем сущность теоремы Шеннона для непрерывного канала с помехами?
21. Какова связь между пропускной способностью и емкостью канала?

Глава VI

ЭФФЕКТИВНОСТЬ ИНФОРМАЦИОННЫХ СИСТЕМ

6.1. Критерии оценки эффективности информационных систем

Эффективность информационной системы характеризует ее способность обеспечить данное количество информации с наименьшими затратами времени, полосы частот и мощности сигнала, затратами на обслуживание и т. д.

Одним из показателей эффективности системы может служить ее быстродействие, т. е. скорость передачи информации.

Для сравнительной оценки эффективности различных систем используется критерий «удельной скорости» передачи информации,

т. е. отношение скорости передачи информации к полосе частот F_c, занимаемой в канале связи

$$R = \frac{\bar{I}_T}{F_c} = \frac{I_T}{F_c T_c}, \qquad (6.1)$$

где T_c — время передачи сигнала; $I_т$ — количество информации, передаваемое за время T_c.

В тех случаях, когда мощность передатчика жестко ограничена (например, в системах космической связи), в качестве критерия эффективности применяется коэффициент использования мощности сигнала

Таблица 6

Вид связи	$\bar{I}$ $\left(\frac{\text{дв. ед.}}{\text{с}}\right)$	ν
Телеграф Морзе	$8 \cdot 10^2$	0,45
Телеграф Бодо	80	0,45
Фототелеграф	$2{,}2 \cdot 10^4$	0,68
Телефон импульсный	$5{,}6 \cdot 10^4$	0,79
Телевидение	$5{,}9 \cdot 10^7$	0,72

$$\beta = \frac{I_T}{\frac{P_у}{P_\xi} F_c T_c}, \qquad (6.2)$$

где $P_у/P_\xi$ — отношение мощности сигнала к мощности помехи в канале передачи информации.

Коэффициент β играет большую роль при сравнительной оценке и энергетическом расчете радиоканалов с очень большими потоками информации. Он показывает, во сколько раз средняя мощность сигнала должна превышать среднюю мощность помех для обеспечения заданной скорости передачи информации на 1 Гц полосы частот, занимаемой в канале связи.

Часто эффективность информационных систем характеризуют коэффициентом

$$\nu = \frac{I_T}{F_c T_c \log_2 \frac{P_у}{P_\xi}}, \qquad (6.3)$$

получившим название удельной содержательности сигнала, определяющим количество информации в единице объема сигнала.

В табл. 6 приведены типовые значения скорости передачи информации и удельной содержательности сигнала для наиболее распространенных видов связи [15].

Как видно из таблицы, наибольшим быстродействием из приведенных в ней систем передачи информации обладает телевидение, но наиболее экономичной системой является телефон.

6.2. Способы повышения эффективности информационных систем

Как видно из выражений (6.1)—(6.3), при наложении ограничений на физические параметры сигнала (мощность, частоту, длительность и т. д.) эффективность систем может быть повышена за

счет увеличения скорости передачи информации, т. е. для повышения эффективности системы передачи информации необходимо повышать энтропию сообщений. Энтропия сообщений зависит от закона распределения вероятностей. Следовательно, для повышения эффективности необходимо осуществить перераспределение плотностей элементов сообщения.

Известно также, что при наличии корреляционных связей между элементами сообщений энтропия последних уменьшается. Поэтому повышение эффективности информационных систем можно также получить за счет устранения или ослабления взаимосвязей между элементами сообщений, т. е. за счет декорреляции сообщений.

Наконец, повышение эффективности систем можно получить за счет соответствующего выбора способа кодирования, обеспечивающего экономию во времени при передаче сообщений. Очевидно, что наибольшую эффективность системы даст код, при котором среднее количество кодовых символов, приходящееся на один элемент сообщения, будет минимальным. Такой код называют эффективным.

Взаимные связи, существующие между отдельными сообщениями, позволяют по данным наблюдений за предыдущими сообщениями предсказывать последующие сообщения. Тогда, вычитая из предсказанного сообщения истинное, можно в линию посылать полученную разность (сигнал ошибки). Разностный сигнал несет по существу те новые сведения, которые не могли быть получены ранее по известным корреляционным связям между сообщениями.

Поскольку среднее значение сигнала ошибки меньше среднего значения сигнала, то такой способ позволяет уменьшить объем сигнала, а следовательно, увеличить эффективность системы.

Выбор того или иного способа повышения эффективности систем из известных должен производиться с учетом сложности его технической реализации, а также с учетом обеспечения необходимой помехоустойчивости систем.

Рассмотрим более детально некоторые наиболее широко известные способы повышения эффективности систем передачи информаций.

6.3. Перераспределение плотностей вероятностей элементов сообщения

Ранее отмечалось, что наибольшей энтропией обладают сообщения при равновероятном и независимом появлении их элементов. Однако вывод был сделан без учета энергетических характеристик сигнала. Более полное рассмотрение этого вопроса показало, что при наложении ограничений на мощность сигнала максимум энтропии дает симметричное нормальное распределение. Поэтому в реальных системах передачи информации целесообразно применять распределение, приближающееся к нормальному. Следовательно, если источник вырабатывает сообщения с распределением элементов, отличным от нормального, то в целях лучшего использования канала

передачи информаций необходимо это распределение изменить так, чтобы оно приблизилось к нормальному закону. В общем случае изменение закона распределения возможно лишь за счет нелинейных преобразований передаваемого сообщения. Сущность такого преобразования рассмотрим на конкретном примере, описанном в [34] и иллюстрируемом на рис. 6.1.

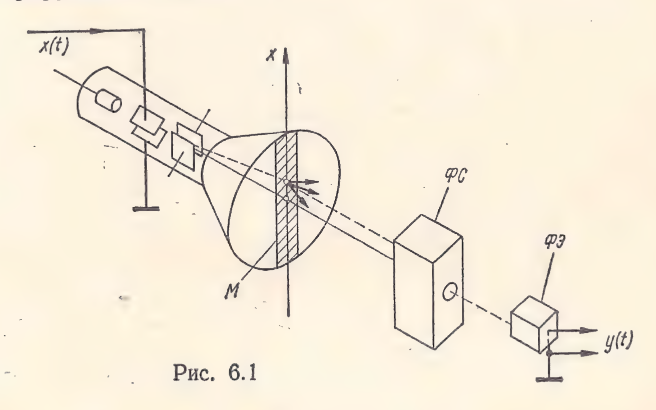


Рис. 6.1

Сигнал x, характеризуемый плотностью распределения вероятностей $w(x)$ и подлежащий преобразованию, подается на отклоняющие пластины электронно-лучевой трубки. Функции нелинейного преобразователя выполняет помещаемая перед экраном трубки маска M с переменной прозрачностью, закон изменения которой описывается нелинейной функцией $f(x)$ (рис. 6.2).

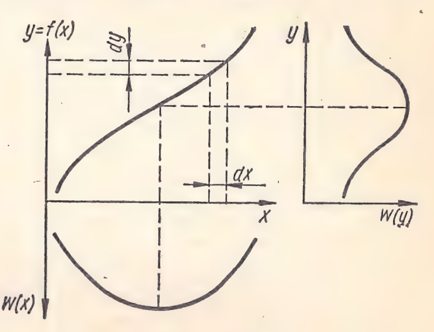


Рис. 6.2

Световой поток Ф, проходящий через маску, находится в нелинейной зависимости от x

$$\Phi = k_1 f(x), \qquad (6.4)$$

где k_1 — коэффициент пропорциональности.

С помощью специальной фокусирующей системы Φ_c поток Ф попадает на фотоэлемент Ф. Э., ток которого пропорционален этому потоку

$$I = k\Phi = k_1 k_2 f(x) = kf(x). \qquad (6.5)$$

Следовательно, если раньше вероятность уровня x определялась плотностью $w(x)$, то теперь та же плотность отвечает уровню $y = f(X)$. Таким образом, ток на выходе фотоэлемента представляет собой новый сигнал с преобразованным распределением вероятностей $w(y)$.

Установим зависимость между плотностями распределения вероятностей входного и выходного сигналов преобразователя.

Из рис. 6.2 видно, что вероятность попадания входного сигнала x в интервал dx равна вероятности попадания выходного сигнала y в интервал dy, т. е.

$$w(y)\,dy = w(x)\,dx, \tag{6.6}$$

где $w(y)$ — плотность распределения вероятностей выходного сигнала y.

Функции (6.5) соответствует обратная функция

$$x = f^{-1}(y), \tag{6.7}$$

тогда

$$dx = \frac{d\,[f^{-1}(y)]}{dy}\,dy. \tag{6.8}$$

Подставляя (6.7) и (6.8) в (6.6), найдем

$$w(y)\,dy = w\,[f^{-1}(y)[\,\frac{d\,[f^{-1}(y)]}{dy}\,dy,$$

откуда

$$w(y) = w\,[f^{-1}(y)]\,\frac{d\,[f^{-1}(y)]}{dy}\,. \tag{6.9}$$

Очевидно, что сигнал на выходе передающего устройства будет искажен вследствие прохождения его через нелинейный четырехполюсник, поэтому в приемном устройстве полученный сигнал необходимо подвергнуть обратной операции.

6.4. Декорреляция сообщений

Известно несколько способов декорреляции сообщений: метод укрупнения, метод предсказания и др. Эти способы подробно изложены в литературе [15, 24, 32, 34, 35].

Рассмотрим в качестве примера сущность метода укрупнения.

Идея декорреляции методом укрупнения заключается в следующем. Сигнал разбивается не на отдельные элементы, а на отрезки (полиграммы), каждый из которых содержит группу элементов. Эти отрезки рассматриваются как элементы нового сигнала. При этом оказывается, что взаимосвязь между укрупненными элементами будет слабее, чем между элементами у исходного сигнала.

Декорреляция сигналов тем значительнее, чем большее количество элементов включено в полиграммы.

При укрупнении сигнала кодированию подвергаются не отдельные элементы, а их группы, т. е. осуществляется переход к коду с более высоким основанием

$$m_2 = m_1^r. \tag{6.10}$$

где m_1 — первоначальное основание кода; r — число элементов в полиграмме; m_2 — новое основание кода.

Энтропия, как известно, обладает важным свойством аддитивности. Переход к укрупненным элементам кода обеспечивает увеличение среднего значения энтропии на элемент сообщения в r раз, так как среднее значение энтропии на полиграмму равно сумме средних значений энтропии на элемент полиграммы. Как при этом изменится избыточность сообщений?

Если среднюю энтропию на элемент исходного сообщения принять равной h_1, то после укрупнения средняя энтропия на элемент будет равна $h_2 = rh_1$. Максимальная энтропия сообщения с укрупненными элементами

$$H_{2\,\text{макс}} = \log_2 N_2 = r \log_2 N_1 = rH_{1\,\text{макс}},$$

где N_2 — объем укрупненного алфавита; N_1 — объем первичного алфавита; $H_{1\,\text{макс}}$ — максимальная энтропия первичного сообщения.

Избыточность сообщения с укрупненными элементами

$$k_{\text{и}2} = 1 - \frac{H_2}{H_{2\,\text{макс}}} = 1 \cdot \frac{H_1}{H_{1\,\text{макс}}} = k_{\text{и}1}, \tag{6.11}$$

где $k_{\text{и}1}$ — избыточность первичного сообщения.

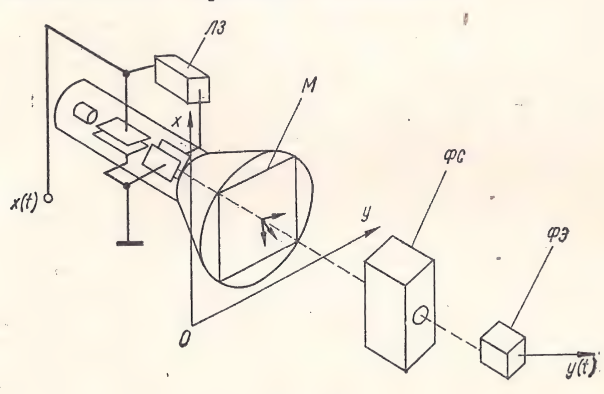


Рис. 6.3

Выражение (6.11) показывает, что укрупнение элементов не обеспечивает уменьшения избыточности сообщений. Это объясняется тем, что при укрупнении элементов уменьшение избыточности за счет ослабления связей между элементами компенсируется увеличением избыточности за счет более неравномерного распределения вероятростей элементов [61]. Поэтому для полного устранения избыточности укрупнение элементов должно сочетаться с использованием оптимального статистического кода для кодирования укрупненного алфавита или других мер, обеспечивающих перераспределение вероятностей элементов сообщений.

В книге [35] описано устройство, совмещающее в себе функции формирования укрупненного сигнала и перераспределения вероятностей. В устройстве учитываются лишь взаимосвязи двух соседних

элементов исходного сигнала. Схема этого устройства изображена на рис. 6.3.

Исходный сигнал подается на вертикально отклоняющие пластины электронно-лучевой трубки, а на горизонтально отклоняющие пластины поступает тот же сигнал, задержанный линией задержки $Л_з$ на время, равное длительности одного элемента. Луч трубки в этом случае отклоняется по горизонтали на величину, пропорциональную значению предшествующего элемента, а по вертикали — на величину, пропорциональную значению данного элемента. Положение луча на экране трубки определяется, таким образом, парой элементов l_i и l_j.

Закон распределения прозрачности маски $М$, расположенной перед экраном трубки, выбран так, чтобы обеспечивался нормальный закон распределения вероятностей светового потока, поступающего через маску и фокусирующую систему ФС на фотоэлемент ФЭ. Следовательно, ток фотоэлемента будет соответствовать укрупненному сигналу с нормальным законом распределения вероятностей.

6.5. Оптимальное статистическое кодирование

Оптимальное статистическое кодирование обеспечивает минимизацию среднего количества кодовых символов на один элемент сообщения. Этим обеспечивается получение максимально возможного количества информации, передаваемого кодовыми комбинациями при заданной длительности работы канала, а следовательно, и пропускной способности канала.

Предположим, что кодирующее устройство может формировать m различных кодовых комбинаций с длительностями $t_1, t_2, \ldots, t_m$.

Задача состоит в том, чтобы установить такую зависимость между длительностями кодовых комбинаций t_i и вероятностями их поступления p_i, при которой обеспечивается максимум скорости передачи информации.

Пусть для передачи сообщения длительностью T необходимо N кодовых комбинаций. Тогда должны быть справедливы следующие равенства:

$$N = \sum_{i=1}^{m} n_i; \tag{6.12}$$

$$T = \sum_{i=1}^{m} n_i t_i; \tag{6.13}$$

$$\sum_{i=1}^{m} p_i = 1; \tag{6.14}$$

$$T = N \sum_{i=1}^{m} p_i t_i, \tag{6.15}$$

где n_i — среднее количество i-х кодовых комбинаций, используемых при передаче сообщений.

Полное количество информации, содержащееся в сообщении длительностью T при независимости появления его элементов

$$I = -N \sum_{i=1}^{m} p_i \log_2 p_i. \qquad (6.16)$$

Поставленная задача сводится к обеспечению максимума количества информации I соответствующим выбором величин N и p_i при соблюдении условий (6.14) и (6.15).

Задача может быть решена методом неопределенных множителей Лагранжа [9]. В соответствии с этим методом составим новую функцию

$$F = -N \sum_{i=1}^{m} p_i \log_2 p_i + \lambda_1 \left(1 - \sum_{i=1}^{m} p_i\right) + \lambda_2 \left(T - N \sum_{i=1}^{m} p_i t_i\right).$$

Определим максимум этой функции, для чего найдем предварительно ее частные производные по N и p_i и приравняем их к нулю

$$\frac{\partial F}{\partial N} = -\sum_{i=1}^{m} p_i \log_2 p_i - \lambda_2 \sum_{i=1}^{m} p_i t_i = 0;$$

$$\frac{\partial F}{\partial p_i} = -N\left(\log_2 p_i + \frac{1}{\ln 2}\right) - \lambda_1 - \lambda_2 N t_i = 0 \quad (i = 1, 2, \ldots, m).$$

Из второго уравнения найдем

$$\log_2 p_i = -\frac{\lambda_1}{N} - \lambda_2 t_i - \frac{1}{\ln 2} \quad (i = 1, 2, \ldots, m) \qquad (6.17)$$

и, подставив это выражение в первое уравнение, будем иметь

$$\sum_{i=1}^{m} p_i \left(-\frac{\lambda_1}{N} - \frac{1}{\ln 2}\right) = 0,$$

откуда

$$\lambda_1 = -\frac{N}{\ln 2}.$$

Подставив найденное значение λ_1 в (6.17), получим

$$\log_2 p_i = -\lambda_2 t_i \quad (i = 1, 2, \ldots, m),$$

откуда

$$p_i = 2^{-\lambda_2 t_i} \quad (i = 1, 2, \ldots, m). \qquad (6.18)$$

Выражение (6.18) показывает характер зависимости между длительностями кодовых комбинаций и вероятностями их появления при оптимальном кодировании. Из выражения видно, что кодовые комбинации с малой вероятностью появления должны иметь большую длительность и наоборот.

Для уяснения физического смысла коэффициента λ_2 подставим (6.18) в (6.16). Так как условие (6.18) определяет условие максимизации количества передаваемой информации за заданное время T, то получим

$$-N\sum_{i=1}^{m} p_i\,(-\lambda_2 t_i) = N\lambda_2 \sum_{i=1}^{m} p_i t_i = \lambda_2 T = I_{\text{макс}},$$

откуда

$$\lambda_2 = \frac{I_{\text{макс}}}{T} = C,$$

т. е. коэффициент λ_2 численно равен пропускной способности канала передачи информации.

Таким образом,

$$p_i = 2^{-Ct_i} \quad (i = 1,\ 2,\ \ldots,\ m), \tag{6.19}$$

или

$$t_i = -\frac{\log_2 p_i}{C} \quad (i = 1,\ 2,\ \ldots,\ m). \tag{6.20}$$

Выражения (6.19) и (6.20) определяют условие согласования работы кодирующего устройства со статистическими свойствами передаваемых сообщений, обеспечивающее передачу информации по каналу со скоростью, практически равной его пропускной способности.

В настоящее время разработано большое количество различных способов оптимального статистического кодирования. Все они должны обеспечивать решение двух основных задач:

1) при заданной статистике источника сообщений формирование кодовых комбинаций со статистическими характеристиками, при которых достигается приближение скорости передачи информации к пропускной способности канала;

2) возможность однозначного декодирования сигналов на приемной стороне.

Для двоичного канала с отсутствием статистических связей между символами этим требованиям удовлетворяет код Шеннона — Фано.

Известно, что при отсутствии статистических связей между символами скорость передачи информации будет максимальна при условии равной вероятности передачи символов 0 и 1. В соответствии с этим построение кода Шеннона — Фано производится методом дихотомий (последовательного деления пополам). Все подлежащие кодированию символы сообщения разбиваются на две группы так, чтобы суммы вероятностей появления элементов сообщений в каждой группе были бы по возможности одинаковы.

В результате такого разбиения как бы образовано новое сообщение, состоящее всего из двух элементов, вероятности появления которых примерно одинаковы. Всем символам первой группы приписывается 0 и всем символам второй группы — 1. Каждая из

полученных групп затем разбивается на две подгруппы с одинаковыми суммарными вероятностями и т. д. Процесс деления повторяется до тех пор, пока в каждой подгруппе останется по одному символу.

Методика построения кода Шеннона — Фано для случая передачи четырех символов сообщения с вероятностями $p(x_1) = 0{,}5$; $p(x_2) = 0{,}25$; $p(x_3) = 0{,}125$ и $p(x_4) = 0{,}125$ иллюстрируется табл. 7. Для удобства построения все символы сообщения выписываются в таблицу в порядке убывания вероятностей. При разбиении верхним половинам групп приписывается символ 0 и нижним — символ 1.

Как видно из табл. 7, полученный код является неравномерным, так как длина кодовых комбинаций находится в обратной зависимости от их вероятности. Для любой позиции всей совокупности кодовых комбинаций вероятности передачи 0 и 1 одинаковы.

Таблица 7

Символы сообщения	Вероятности	Этапы деления на подгруппы				Символы кодовых комбинаций			
		1	2	3	4	1	2	3	4
x_1	0,5	} I				0			
x_2	0,25	} II	} I			1	0		
x_3	0,125		} II	} I		1	1	0	
x_4	0,125			} II		1	1	1	

Помимо того, из таблицы видно, что ни одна из кодовых комбинаций не является началом другой. Этим обеспечивается требование разделимости кодовых комбинаций, т. е. возможность однозначного декодирования сигналов.

Подсчитаем скорость передачи информации, которая обеспечивается полученным кодом.

Пусть длительность символов кодовых комбинаций равна τ. Тогда средняя длительность кодовых комбинаций

$$\overline{T} = p(x_i) \cdot \tau + p(x_2) \cdot 2\tau + p(x_3) \cdot 3\tau + p(x_4)\, 3\tau = 1{,}75\tau.$$

Средняя энтропия на символ сообщения

$$H(X) = -\sum_{i=1}^{4} p(x_i) \log_2 p(x_i) = 1{,}75 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Таким образом, скорость передачи информации

$$\overline{I}(X) = \frac{1}{\tau} = C \ \frac{\text{дв. ед.}}{\text{с}}.$$

Следовательно, полученный код в рассматриваемом случае позволил получить максимально возможное значение скорости передачи информации, т. е. обеспечить полное согласование статистических характеристик источника сообщений со свойствами канала.

Это удалось благодаря тому, что в рассмотренном примере значения вероятностей $p(x_i)$ выбраны такими, что условия деления на подгруппы удается выполнить точно. В реальных условиях это, как правило, не обеспечивается и скорость передачи информации будет меньше пропускной способности канала.

Эффективность кодирования может быть при необходимости увеличена путем перехода от кодирования одиночных символов сообщения к кодированию групп символов сообщения, причем с укрупнением групп эффективность будет повышаться. Повышение эффективности происходит при этом за счет того, что при укрупнении групп получающийся набор вероятностей можно делить на более близкие по суммарной вероятности подгруппы. Методика построения кода Шеннона — Фано путем кодирования групп символов сообщения иллюстрируется в 6.6 примером 6.4.

6.6. Примеры

Пример 6.1 (из [26]). Плотность распределения вероятностей сигнала описывается законом

$$w(x) = \begin{cases} \exp\{-x\} & \text{при } x \geqslant 0; \\ 0 & \text{при } x < 0. \end{cases}$$

Сигнал пропускается через нелинейный преобразователь с характеристикой вида

$$y = f(x) = \left|\sqrt[4]{x}\right|.$$

Найти плотность распределения вероятностей выходного сигнала преобразователя.

Решение. Для определения закона распределения выходного сигнала воспользуемся соотношением (6.9).

Обратная функция преобразователя

$$x = f^{-1}(y) = y^4,$$

производная этой функции

$$\frac{d\,[f^{-1}(y)]}{dy} = 4y^3.$$

Тогда искомая функция распределения

$$w(y) = w\,[f^{-1}(y)]\,\frac{d\,[f^{-1}(y)]}{dy}\,4\exp\{-y^4\}\,y^3.$$

Пример 6.2. По каналу необходимо передавать сообщение, формируемое из независимых символов с вероятностями появления, определяемыми табл. 8.

Канал рассчитан на передачу элементов сообщения со средней длительностью $\bar{\tau} = 1$ мс.

Требуется построить оптимальный код с время-импульсной модуляцией.

Решение. Код будем строить на основании зависимости (6.16).

Пропускную способность канала определяем из отношения (5.12).

Средняя скорость передачи символов сигнала по условию задачи

Таблица 8

Символы	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Вероятность	0,25	0,20	0,15	0,15	0,10	0,05	0,05	0,03	0,01	0,01

$$\overline{U}_y = \frac{1}{\tau} = 1000 \ \frac{\text{симв.}}{\text{с}}.$$

Тогда пропускная способность канала

$$C = \overline{U}_y \log_2 n = 1000 \log_2 10 = 3320 \ \frac{\text{дв. ед.}}{\text{с}}.$$

Данные расчета оптимального кода сведены в табл. 9.

Таблица 9

Символы	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$-\log_2 p(x_i)$	2,00	2,32	2,74	2,74	3,32	4,32	4,32	5,06	6,64	6,64
l_i, мс	0,602	0,698	0,827	0,827	1,0	1,3	1,3	1,52	2,0	2,0

Пример 6.4 (из [10]). По каналу необходимо передать сообщение, формируемое из 3-х независимых символов, с вероятностями появления, определяемыми табл. 10.

Таблица 10

Символы	x_1	x_2	x_3
Вероятность	0,2	0,7	0,1

Таблица 11

Символ сообщения	Код
x_1	00
x_2	01
x_3	10

Канал обладает полосой пропускания, допускающей передачу кодовых комбинаций с длительностью символов $\tau = 1$ мс.

Определить скорость передачи информации по каналу при использовании:

а) равномерного двоичного кода;

б) кода Шеннона — Фано при кодировании отдельных символов сообщений;

в) кода Шеннона—Фано при кодировании групп из двух символов сообщения.

Решение. а) длина (количество разрядов) n равномерного двоичного кода определяется из отношения

$$n = \log_2 m,$$

где m — число элементов сообщения.

При $m = 3$ длина кода должна быть равна $n = 2$.

Возможные варианты кодовых комбинаций представлены в таблице 11.

Длительности кодовых комбинаций

$$T = 2\tau = 2 \text{ мс}.$$

Энтропия сообщения

$$H(X) = -\sum_{i=1}^{3} p(x_i) \log_2 p(x_i) =$$

$$= -0{,}2 \log_2 0{,}2 - 0{,}7 \log_2 0{,}7 - 0{,}1 \log_2 0{,}1 = 1{,}16 \ \frac{\text{дв. ед.}}{\text{симв.}}.$$

Скорость передачи информации при использовании равномерного двоичного кода

$$\bar{I}(X) = \frac{H(X)}{T} = \frac{1{,}16}{2 \cdot 10^{-3}} = 580 \ \frac{\text{дв. ед.}}{\text{с}};$$

б) построение кода Шеннона — Фано с кодированием отдельных символов сообщения производится в соответствии с методикой, изложенной в 6.5, и иллюстрируется табл. 12.

Таблица 12

Символ сообщений	Вероятности	Деление на подгруппы		Код	Длительность кодовых комбинаций
x_1	0,7	I		0	τ
x_2	0,2	II	I	10	2τ
x_3	0,1		II	11	2τ

Средняя длительность кодовых комбинаций при этом

$$\bar{T} = \sum_{i=1}^{3} p(x_i) T_i = (0{,}7 \cdot 10^{-3} + 0{,}2 \cdot 2 \cdot 10^{-3} +$$

$$+ 0{,}1 \cdot 2 \cdot 10^{-3}) = 1300 \text{ мкс}.$$

Скорость передачи информации

$$\bar{I}(X) = \frac{1{,}16}{1300 \cdot 10^{-6}} = 890 \ \frac{\text{дв. ед.}}{\text{с}};$$

в) для построения кода Шеннона — Фано с кодированием групп из двух символов сообщения вычисляем вероятности отдельных групп по формуле $p(x_i x_j) = p(x_i)\,p(x_j)$, рассматриваем в дальнейшем полученные группы как символы нового сообщения и производим кодирование по методике пункта (б). Построение кода иллюстрируется табл. 13.

Таблица 13

Группы символов	Вероятность	Деление на подгруппы						Код	Длительность кодовой комбинации
x_2x_2	0,49	I						0	τ
x_1x_2	0,14		I	I				100	3τ
x_2x_1	0,14			II				101	3τ
x_2x_3	0,07			I	I			1100	4τ
x_3x_2	0,07				II			1101	4τ
x_1x_1	0,04	II			I			1110	4τ
x_1x_3	0,02		II	II		I		11110	5τ
x_3x_1	0,02				II	II	I	111110	6τ
x_3x_3	0,01						II	111111	6τ

Средняя длительность кодовых комбинаций

$$\bar{T} = 0{,}49\tau + 2 \cdot 0{,}14 \cdot 3\tau + 2 \cdot 0{,}07 \cdot 4\tau + 0{,}04 \cdot 4\tau + \\ + 0{,}02 \cdot 5\tau + 0{,}02 \cdot 6\tau + 0{,}01 \cdot 6\tau = 1{,}165 \text{ мс}.$$

Скорость передачи информации

$$\bar{I}(X) = \frac{1{,}16}{1{,}165 \cdot 10^{-3}} = 995 \ \frac{\text{дв. ед.}}{\text{с}}.$$

Пропускная способность двоичного канала

$$C = \frac{1}{\tau} = 1000 \ \frac{\text{дв. ед.}}{\text{с}}.$$

Таким образом, в случае (в) обеспечивается наибольшее приближение скорости передачи информации к пропускной способности канала.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под эффективностью системы передачи информации и какими показателями она характеризуется?
2. Как зависит эффективность системы передачи информации от избыточности сообщений?
3. В чем сущность метода повышения эффективности систем за счет перераспределения плотностей вероятностей элементов сообщения?
4. Как влияет корреляция между элементами сообщений на эффективность системы передачи информации?
5. В чем состоит суть метода укрупнения?
6. В чем состоит суть оптимального статистического кодирования?
7. Каким основным требованиям должны удовлетворять оптимальные статистические коды?
8. Объясните методику построения кода Шеннона — Фано.

Глава VII

ПОМЕХОУСТОЙЧИВОСТЬ ИНФОРМАЦИОННЫХ СИСТЕМ

7.1. Общая характеристика помех в системах передачи информации

Под помехами понимаются любые возмущения в канале передачи информации, вызывающие случайные отклонения принятого сообщения от переданного [2]. Помехи обычно классифицируются по месту их возникновения, по статистическим свойствам и по характеру воздействия на полезный сигнал.

По месту возникновения помехи можно разделить на внешние и внутренние. К внешним помехам относятся помехи, источники которых находятся вне системы передачи информации. Сюда можно отнести:

1) атмосферные помехи (вызванные грозовыми разрядами);

2) космические помехи, вызванные радиоизлучением Солнца и других небесных тел;

3) промышленные помехи, обусловленные работой различных электрических устройств и агрегатов.

Внутренние помехи возникают в самой аппаратуре системы передачи информации. Сюда можно отнести помехи в виде тепловых шумов электронных ламп, полупроводниковых приборов, сопротивлений и других элементов; помехи, вызванные изменением параметров линий связи, влиянием линий друг на друга, а также за счет кратковременных разрывов связи; помехи, возникающие при преобразовании сигналов в отдельных элементах системы (шумы квантования, искажения сигналов за счет ограниченного значения полосы пропускания элементов, за счет нелинейности характеристик преобразования и пр.); помехи, обусловленные нестабильностью элементов аппаратуры, а также аппаратурные искажения, вызванные технической неисправностью или недостаточно точной настройкой аппаратуры.

По своим свойствам помехи могут быть детерминированными и случайными. Защита против детерминированных помех не вызывает особых затруднений. В дальнейшем рассмотрим только случайные помехи.

Все случайные помехи можно объединить в три группы:

1) импульсные помехи;

2) флюктуационные помехи;

3) синусоидальные помехи.

Импульсные помехи представляют в общем случае последовательность импульсов произвольной формы со случайными амплитудой, длительностью и моментом появления. Характерной особенностью импульсных помех является то, что переходные процессы, вызванные в аппаратуре каким-либо импульсом, успевают практически затухнуть до появления следующего импульса.

Характерными примерами импульсных помех являются помехи от грозовых разрядов, от системы зажигания двигателей внутреннего сгорания, помехи, связанные с коммутационными процессами и т. п.

Флюктуационная помеха представляет собой совокупность большого числа кратковременных нерегулярных импульсов со случайными параметрами. Переходные процессы от воздействия отдельных импульсов, накладываясь друг на друга, образуют непрерывный случайный процесс. Характерной особенностью этих помех является отсутствие выбросов, превышающих средний уровень более чем в три-четыре раза [20, 36].

Так как длительность переходного процесса определяется полосой пропускания канала передачи информации, то и характер помех зависит от ширины полосы канала. Одна и та же помеха может быть импульсной для широкополосной и флюктуационной для узкополосной системы.

Флюктуационные помехи представляют собой обычно белый шум, гауссов шум или белый гауссов шум. Последний характерен как своей распространенностью, так и тем, что он принципиально не может быть устранен.

К таким помехам можно отнести:

1) тепловые шумы сопротивлений и полупроводниковых приборов, дробовый эффект электронных ламп;

2) космические помехи;

3) атмосферные помехи в диапазоне коротких волн и пр.

Синусоидальные помехи представляют собой синусоидальные колебания со случайно изменяющимися амплитудой, фазой и частотой. Эти помехи характеризуются медленным изменением параметров, вследствие чего ширина спектра модулирующей функции синусоидальной помехи оказывается практически малой по сравнению с полосой пропускания канала.

В качестве источников синусоидальных помех могут быть посторонние радиоустановки, генераторы переменного тока и пр.

По характеру воздействия на полезный сигнал помехи подразделяются на аддитивные и мультипликативные. Аддитивная помеха — это помеха, представляемая не зависящим от сигнала случайным слагаемым. Аддитивную помеху называют иногда «шумом». Мультипликативная помеха — это помеха, представляемая не зависящим от сигнала случайным множителем [2]. Подавляющая часть встречающихся на практике помех принадлежит к группе аддитивных помех.

Характерным примером мультипликативной помехи является искажение сигнала за счет случайных изменений характеристик канала передачи информации.

Все случайные помехи представляют собой случайный процесс и описываются с помощью функций распределения вероятностей либо числовых характеристик в виде моментов распределения.

7.2. Критерии оценки помехоустойчивости информационных систем

Помехоустойчивость определена в первой главе как способность информационной системы противостоять вредному действию помех. В результате действия помех принятое сообщение будет в какой-то мере отличаться от переданного. Поэтому помехоустойчивость можно характеризовать как степень соответствия принятого сообщения переданному при заданной помехе. При сравнении нескольких систем та из них будет более помехоустойчивой, которая при одинаковой помехе обеспечит меньшее различие между принятым и переданным сообщениями.

Для характеристики степени соответствия принятого сообщения переданному введен термин «верность» [37]. Количественная мера этого соответствия выбирается по-разному в зависимости от характера сообщений.

При передаче непрерывных сообщений часто в качестве критерия верности используется критерий среднеквадратического отклонения принятого сообщения $Y(t)$ относительно переданного $X(t)$

$$\sigma = \sqrt{\overline{|Y(t) - X(t)|^2}}. \tag{7.1}$$

Применяются также критерий абсолютного отклонения

$$\delta_{\text{абс}} = \overline{|Y(t) - X(t)|} \tag{7.2}$$

и критерий наибольшего отклонения

$$\delta_{\text{макс}} = \max |Y(t) - X(t)|. \tag{7.3}$$

Обычно для удобства сравнительной оценки помехоустойчивости различных систем рассматриваются относительные отклонения

$$\gamma_\sigma = \frac{\sigma}{X_э}; \quad \gamma_{\text{абс}} = \frac{\delta_{\text{абс}}}{X_э}; \quad \gamma_{\text{макс}} = \frac{\delta_{\text{макс}}}{X_э} \tag{7.4}$$

или приведенные отклонения

$$\alpha_\sigma = \frac{\sigma}{L_x}; \quad \alpha_{\text{абс}} = \frac{\delta_{\text{абс}}}{L_x}; \quad \alpha_{\text{макс}} = \frac{\delta_{\text{макс}}}{L_x}, \tag{7.5}$$

где $X_э$ — эффективное значение сообщения; L_x — динамический диапазон передаваемых сообщений.

Если предположить, что канал передачи информации имеет идеальную П-образную амплитудно-частотную и линейную фазочастотную характеристики, то при наличии флюктуационной помехи типа белого гауссова шума среднеквадратическое значение отклонения принятого сообщения относительно переданного равно корню квадратному из средней мощности помехи на выходе приемника $P_{\text{п. вых}}$, а относительная величина этого отклонения выражается

через корень квадратный отношения средних мощностей помехи и сигнала на выходе приемника

$$\gamma_\sigma = \sqrt{\left(\frac{P_\xi}{P_x}\right)_{\text{вых}}}. \qquad (7.6)$$

Для сравнительной оценки систем и практических расчетов часто в качестве критерия помехоустойчивости принимают величину «выигрыша системы»

$$B = \frac{\left(\frac{P_x}{P_\xi}\right)_{\text{вых}}}{\left(\frac{P_x}{P_\xi}\right)_{\text{вх}}}, \qquad (7.7)$$

где $\left(\frac{P_x}{P_\xi}\right)_{\text{вых}}$ и $\left(\frac{P_x}{P_\xi}\right)_{\text{вх}}$ — отношение средних мощностей сигнала и помехи на выходе и входе устройства.

В случае передачи дискретных сообщений, а также непрерывных сообщений с кодоимпульсной модуляцией сигналов, в качестве критерия верности целесообразно использовать вероятность правильного приема

$$p_{\text{пр}} = 1 - p_{\text{ош}}, \qquad (7.8)$$

где $p_{\text{ош}}$ — вероятность ошибки в воспроизведении сообщения.

В реальных условиях вероятность ошибки $p_{\text{ош}}$ очень мала и значительно меньше единицы. Поэтому очень часто используют для оценки помехоустойчивости логарифмическую величину

$$S = \lg\frac{1}{p_{\text{ош}}} = \lg\frac{1}{1 - p_{\text{пр}}}. \qquad (7.9)$$

По смыслу определения помехоустойчивость понимается как свойство системы передачи информации в целом. Однако оценка помехоустойчивости системы в целом — довольно сложная задача. Поэтому обычно говорят о помехоустойчивости отдельных звеньев системы: о помехоустойчивости передачи (в частности, о помехоустойчивости кода или вида модуляции) и о помехоустойчивости приема. Помехоустойчивость кода может быть оценена величиной (7.9), где $p_{\text{ош}}$ — вероятность искажения кодовой комбинации под воздействием помех.

При оценке влияния вида модуляции на помехоустойчивость системы производят обычно сравнение различных видов модуляции с амплитудой. С этой целью часто применяется коэффициент

$$R_{\text{м}} = \frac{\left(\frac{X_э}{\sigma_\xi}\right)_{\text{вых}}}{\left(\frac{X_э}{\sigma_\xi}\right)_{\text{вых АМ}}}, \qquad (7.10)$$

или

$$R_м = \frac{\left(\frac{P_x}{P_\xi}\right)_{вых}}{\left(\frac{P_x}{P_\xi}\right)_{вых\ АМ}}, \tag{7.11}$$

где $\left(\frac{X}{\sigma_\xi}\right)_{вых}$ и $\left(\frac{P_x}{P_\xi}\right)_{вых}$ — отношение эффективного значения сигнала к среднеквадратичному значению помехи и отношение средней мощности сигнала к средней мощности помехи на выходе приемного устройства при произвольном виде модуляции; $\left(\frac{X_э}{\sigma_\xi}\right)_{вых\ АМ}$ и $\left(\frac{P_x}{P_\xi}\right)_{вых\ АМ}$ — аналогичные отношения на выходе приемного устройства при амплитудной модуляции.

При приеме в зависимости от назначения сигналов могут иметь место три вида задач [37]: 1. Обнаружение сигнала. 2. Различение сигналов. 3. Восстановление сообщений.

Задача обнаружения состоит в том, чтобы по результатам обработки принятого сигнала, который может быть либо только помехой, либо суммой полезного сигнала и помехи, решить, содержится ли полезный сигнал в принятом или нет. При этом могут быть ошибки двоякого рода: 1) при отсутствии полезного сигнала принимается ложное решение о наличии сигнала; 2) при наличии сигнала принимается ложное решение об отсутствии сигнала. Первая ошибка называется ошибкой первого рода или «ложной тревогой». Вторая ошибка называется ошибкой второго рода или «пропуском сигнала». Количественно ошибки первого и второго рода оцениваются условными вероятностями α и β ошибочных решений о наличии полезного сигнала, когда в действительности он отсутствует, и об отсутствии сигнала, когда в действительности он имеется.

Полная вероятность ошибочного решения определяется выражением

$$p_{ош} = q\alpha + p\beta, \tag{7.12}$$

где q и p — априорные вероятности отсутствия и присутствия полезного сигнала.

Помехоустойчивость приемника, решающего задачу обнаружения сигнала, может быть оценена с помощью выражения (7.9), где $p_{ош}$ определяется формулой (7.12).

Когда априори известно, что передан один из двух сигналов x_1 и x_2, то ставится задача различения двух сигналов, т. е. задача определения, имеется ли на входе приемника сигнал x_1 плюс помеха, или сигнал x_2 плюс помеха. Очевидно, помехоустойчивость приемника в этом случае также определяется выражениями (7.9) и (7.12), причем величины α и β в формуле (7.12) являются условными вероятностями ложных заключений о наличии сигналов x_1 или x_2, когда в действительности на вход приемника поступает

сигнал x_2 или x_1 соответственно, а q и p являются априорными вероятностями поступления сигналов x_1 и x_2.

Случай различения многих сигналов в принципиальном отношении мало отличается от случая различения двух сигналов. Помехоустойчивость приемника в этом случае также может быть оценена формулой (7.9), причем вероятность ошибки $p_{\text{ош}}$ определяется с помощью выражения

$$p_{\text{ош}} = \sum_{i=1}^{n} p_i \alpha_i, \tag{7.13}$$

где p_i — априорная вероятность поступления i-го сигнала; α_i — условная вероятность ложного заключения о наличии i-го сигнала, когда в действительности на вход приемника поступает любой другой из n сигналов.

Задача восстановления сообщения значительно отличается от задач обнаружения и различения сигналов. Она сводится к получению выходного сообщения $Y(t)$, наименее отличающегося от передаваемого с точки зрения выбранного критерия верности. Помехоустойчивость приемника в этом случае может быть оценена с помощью критериев отклонения, определяемых выражениями (7.1)—(7.6).

Помехоустойчивость такого приемника может быть также оценена с помощью критерия (7.7). Однако нужно иметь в виду, что эта величина не всегда может быть определена однозначно и не позволяет объективно сравнивать между собой различные системы, если помеха имеет более широкий спектр, чем сигнал и, в частности, если она типа белого шума. Мощность помехи на входе приемника в этом случае определяется полосой пропускания канала связи. Различные системы могут существенно отличаться полосой пропускания каналов связи. При одном и том же канале связи могут применяться приемники с различными полосами пропускания.

Для устранения имеющейся неоднозначности можно сравнивать на входе и выходе приемника отношения мощностей сигнала не с мощностью помехи, а с удельной мощностью помех. В этом случае выигрыш приемника оценивается соотношением

$$B' = \frac{\dfrac{P_x}{P'_{\xi\ \text{вых}}}}{\dfrac{P_x}{P'_{\xi\ \text{вх}}}}, \tag{7.14}$$

где $P'_{\xi\ \text{вых}} = \dfrac{P_{\xi\ \text{вых}}}{F_{\text{вых}}}$ — удельная мощность помех на выходе приемника; $P'_{\xi\ \text{вх}} = \dfrac{P_{\xi\ \text{вх}}}{F_{\text{вх}}}$ — удельная мощность помехи на входе приемника; $F_{\text{вх}}$ и $F_{\text{вых}}$ — полосы частот, в которых измеряется мощность помехи на входе и выходе приемника.

Очевидно,

$$B' = \frac{\beta}{\rho}, \qquad (7.15)$$

где

$$\rho = \frac{F_{\text{вх}}}{F_{\text{вых}}}.$$

7.3. Способы повышения помехоустойчивости информационных систем

В основе всех способов повышения помехоустойчивости информационных систем лежит использование определенных различий между полезным сигналом и помехой. Поэтому для борьбы с помехами необходимы априорные сведения о свойствах помехи и сигнала.

В настоящее время известно большое число способов повышения помехоустойчивости систем. Все эти способы удобнее разбить на две группы. Первая группа способов основана на выборе метода передачи сообщений. Вторая группа способов связана с построением помехоустойчивых приемников.

Простым и часто применяемым способом повышения помехоустойчивости передачи является увеличение отношения сигнал/помеха за счет увеличения мощности передатчика. Однако этот метод, несмотря на свою простоту, может оказаться экономически невыгодным, так как связан с существенным ростом сложности и стоимости оборудования. Помимо того, увеличение мощности передачи сопровождается усилением мешающего действия данного канала на другие.

Важным способом повышения помехоустойчивости передачи непрерывных сигналов является рациональный выбор вида модуляции сигналов. Применяя виды модуляции, обеспечивающие значительное расширение полосы частот сигнала, можно добиться существенного повышения помехоустойчивости передачи. Оценка помехоустойчивости отдельных видов модуляции произведена в параграфе 7.4.

Радикальным способом повышения помехоустойчивости передачи дискретных сигналов является использование специальных помехоустойчивых кодов. При этом имеются два пути повышения помехоустойчивости кодов. Первый заключается в выборе таких способов передачи, которые обеспечивают меньшую вероятность искажения кода, второй — в увеличении корректирующих свойств кодовых комбинаций.

Исследования первого пути показали высокую помехоустойчивость кодов с большим числом частотных признаков [20, 36, 38, 39].

Второй путь связан с использованием кодов, позволяющих обнаруживать и устранять искажения в кодовых комбинациях. Такой способ кодирования связан с введением в код дополнительных, избыточных символов, что сопровождается увеличением времени

передачи либо частоты передачи символов кода. Это приводит к расширению спектра сигнала.

Основные положения теории помехоустойчивого кодирования изложены в параграфе 7.5.

Повышение помехоустойчивости передачи может быть также достигнуто путем повторной передачи одного и того же сообщения. На приемной стороне сравниваются полученные сообщения и в качестве истинных принимаются те, которые имеют наибольшее число совпадений. Чтобы исключить неопределенность при обработке принятой информации и обеспечить отбор по критерию большинства, сообщение должно повторяться не менее трех раз. Очевидно, что этот способ повышения помехоустойчивости связан с увеличением времени передачи.

Системы с повторением передачи дискретной информации делятся на системы с групповым суммированием, у которых сравнение производится по кодовым комбинациям (группам), и на системы с посимвольным суммированием, у которых сравнение осуществляется по символам кодовых комбинаций. Исследования показали, что посимвольная проверка является более эффективной, чем групповая (см. пример 7.5).

Разновидностью систем, у которых повышение помехоустойчивости достигается за счет увеличения времени передачи, являются системы с обратной связью.

При наличии искажений в передаваемых сообщениях информация, поступающая по обратному каналу, обеспечивает повторение передачи. Наличие обратного канала приводит к усложнению системы. Однако в отличие от систем с повторением передачи в системах с обратной связью повторение передачи будет иметь место лишь в случае обнаружения искажений в передаваемом сигнале, т. е. избыточность в целом оказывается меньшей.

Характеристика систем с обратной связью приведена в параграфе 7.13.

Помехоустойчивый прием состоит в использовании избыточности, а также априорных сведений о сигналах и помехах для решения оптимальным способом задачи приема обнаружения сигнала, различения сигналов или восстановления сообщений. В настоящее время для синтеза оптимальных приемников широко используется аппарат теории статистических решений (см. параграфы 8.6—8.11).

Ошибки приемника уменьшаются с увеличением отношения сигнал/помеха на входе приемника. В связи с этим часто производят предварительную обработку принятого сигнала с целью увеличения отношения полезной составляющей к помехе. К таким методам предварительной обработки сигналов относится метод ШОУ (сочетание широкополосного усилителя, ограничителя и узкополосного усилителя), селекция сигналов по длительности, метод компенсации помехи, метод фильтрации, корреляционный метод, метод накопления и др. Описание сущности некоторых из этих методов и их характеристика будут приведены позже (см. параграфы 8.2—8.5).

7.4. Помехоустойчивость различных видов модуляции

Переносчиком информации, как известно, является сигнал, который в общем случае может быть представлен в виде

$$x(t) = f(a_1, a_2, \ldots, a_n, t),$$

где $a_1, a_2, \ldots, a_n$ — параметры сигнала.

Модуляция сигнала состоит в том, что один или несколько его параметров изменяются в соответствии с передаваемым сообщением. Воздействие помехи на носитель приводит к дополнительной (паразитной) модуляции его параметров.

При наложении помехи $\xi(t)$ на полезный сигнал $x(t)$ получается сложный сигнал

$$y(t) = x(t) + \xi(t) + f_1(a_1 + \delta a_1, a_2 + \delta a_2, \ldots, a_n + \delta a_n, t),$$

где $\delta a_1, \delta a_2, \ldots, \delta a_n$ — приращения параметров сигнала под воздействием помехи (паразитная модуляция).

Очевидно, что различные параметры сигнала будут по-разному реагировать на воздействие помехи, т. е. помехоустойчивость различных видов модуляции должна быть неодинаковой. Оценку помехоустойчивости модуляции будем производить с помощью критерия (7.7) и (7.11).

Амплитудная и частотная модуляции

Предположим, что передаваемый сигнал и помеха изменяются по гармоническому закону

$$x(t) = A \sin \omega_0 t;$$
$$\xi(t) = B \sin \omega t.$$

Суммарный сигнал, получаемый в результате наложения помехи на полезный сигнал, можно представить в виде

$$y(t) = A \sin \omega_0 t + B \sin \omega t = (A + \delta A) \sin(\omega_0 + \delta\omega_0) t,$$

где δA и $\delta\omega$ — паразитная модуляция амплитуды и частоты сигнала.

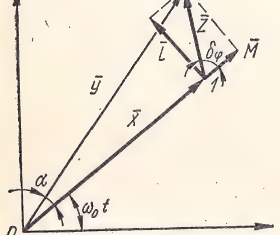


Рис. 7.1

На рис. 7.1 представлена векторная диаграмма, иллюстрирующая образование суммарного сигнала. На этом рисунке $\overline{X}$ — вектор сигнала, вращающийся вокруг точки O с частотой ω_0; $\overline{Z}$ — вектор помехи, вращающийся вокруг точки 1 с частотой ω; $\overline{Y}$ — вектор результирующего сигнала, вращающийся вокруг точки O.

Вектор $\overline{Y}$ совершает сложное движение с периодически изменяющейся скоростью. При этом также будет периодически изменяться

его амплитуда. Периодические изменения скорости вращения вектора $\overline{Y}$ и его амплитуды характеризуют частотную и амплитудную паразитные модуляции сигнала.

Для оценки паразитной модуляции разложим вектор $\overline{Z}$ на составляющие $\overline{M}$ и $\overline{L}$ (рис. 7.1).

Проекция вектора $\overline{Z}$ на вектор $\overline{X}$ выражает амплитудную паразитную модуляцию. Эта проекция

$$M = B\cos(\omega_0 - \omega)t = B\cos\delta\varphi = \delta A \leqslant B, \qquad (7.16)$$

где

$$\delta\varphi = (\omega_0 - \omega)t = \delta_1\omega t.$$

Чтобы найти паразитную частотную модуляцию, необходимо определить угловую скорость вращения вектора $\overline{Y}$. Последняя равна

$$\omega = \frac{d\varphi}{dt},$$

где $\varphi = \omega_0 t + \alpha$ — мгновенная фаза вектора $\overline{Y}$.

Таким образом,

$$\omega' = \omega_0 + \frac{d\alpha}{dt} = \omega_0 + \delta\omega,$$

где $\delta\omega = \frac{d\alpha}{dt}$ — паразитная частотная модуляция.

При $M \ll A$

$$\alpha = \operatorname{tg}\alpha = \frac{B\sin\delta\varphi}{A}.$$

Следовательно,

$$\delta\omega = \frac{d\alpha}{dt} = \frac{B}{A}\cos\delta\varphi\frac{d(\delta\varphi)}{dt} = \frac{B}{A}\delta_1\omega\cos\delta\varphi, \qquad (7.17)$$

где

$$\delta_1\omega = \frac{d(\delta\varphi)}{dt}.$$

Окончательно получим следующее выражение для угловой скорости вектора $\overline{Y}$:

$$\omega' = \omega_0 + \frac{B}{A}\delta\omega\cos\delta\varphi. \qquad (7.18)$$

Оценим величину отношения сигнал/помеха на выходе приемника при амплитудной модуляции. Так как на выходе приемника стоит амплитудный детектор, то выходной сигнал приемника пропорционален изменениям модулируемого параметра.

Следовательно, отношение мощностей сигнала и помехи на выходе приемника при амплитудной модуляции

$$\gamma_{\text{вых } AM} = \frac{(\Delta A)^2}{(\delta A)^2} = \frac{(\Delta A)^2}{M^2_{\text{макс}}},$$

где ΔA — максимальное значение изменения амплитуды сигнала

при полезной модуляции; $M_{макс}$ — максимальное значение изменения амплитуды сигнала при паразитной модуляции.

При стопроцентной амплитудной модуляции

$$\Delta A = A.$$

Тогда

$$\gamma_{вых} = \frac{A^2}{B^2} = \gamma_{вх\ AM},$$

где $\gamma_{вх\ AM} = \frac{A^2}{B^2}$ — отношение мощности полезного сигнала к мощности помехи на входе приемника.

Таким образом, выигрыш амплитудной модуляции

$$B_{AM} = \frac{\gamma_{вых\ AM}}{\gamma_{вх\ AM}} = 1.$$

Следовательно, при амплитудной модуляции отношения сигнал/помеха на выходе и входе приемника одинаковы.

При частотной модуляции используются приемники с частотными детекторами. При этом величина сигнала на выходе приемника будет пропорциональна девиации частоты входного сигнала. Тогда отношение мощностей сигнала и помехи на выходе приемника

$$\gamma_{вых\ ЧМ} = \frac{(\Delta\omega^2)}{(\delta\omega)^2_{макс}} = \frac{A^2(\Delta\omega)^2}{B^2(\delta_1\omega)^2} = \left(\frac{\Delta\omega}{\delta_1\omega}\right)^2 \gamma_{вх\ ЧМ},$$

где $\Delta\omega$ и $(\delta\omega)_{макс}$ — максимальные изменения частоты сигнала при полезной и паразитной частотной модуляции.

Следовательно, выигрыш, обеспечиваемый частотной модуляцией,

$$B_{ЧМ} = \frac{\gamma_{вых\ ЧМ}}{\gamma_{вх\ ЧМ}} = \left(\frac{\Delta\omega}{\delta_1\omega}\right)^2. \qquad (7.19)$$

Обычно $\Delta\omega > \delta_1\omega$, поэтому частотная модуляция подавляет помеху и это подавление будет тем эффективнее, чем больше девиация частоты $\Delta\omega$ при полезной модуляции. Таким образом, частотная модуляция по сравнению с амплитудной обеспечивает выигрыш в $\left(\frac{\Delta\omega}{\delta_1\omega}\right)^2$ раз.

Общий случай — по каналу передается произвольный сигнал. Пусть сигнал характеризуется энергетической спектральной плотностью $G(\omega)$. В канале действует аддитивная помеха типа белый шум с удельной энергетической спектральной плотностью P_0.

Отношение мощностей сигнала и помехи на выходе приемника для случая амплитудной модуляции

$$\gamma_{вых\ AM} = \frac{\int_0^{\Omega} G(\omega)\, d\omega}{P_0\Omega}, \qquad (7.20)$$

где Ω — полоса частот, в которой действуют сигнал и помеха.

Это же отношение для случая частотной модуляции

$$\gamma_{\text{вых } ЧМ} = \frac{\int\limits_0^{\Omega} (\Delta\omega)^2 G(\omega)\, d\omega}{\int\limits_{\omega_0}^{\omega_0+\Omega} (\omega - \omega_0)^2 P_0\, d\omega} = \frac{(\Delta\omega)^2 \int\limits_0^{\Omega} G(\omega)\, d\omega}{P_0 \int\limits_0^{\Omega} (\delta_1\omega)^2\, d(\delta_1\omega)} =$$

$$= \frac{3(\Delta\omega)^2 \int\limits_0^{\Omega} G(\omega) d\omega}{P_0\Omega^3} = \frac{3(\Delta\omega)^2}{\Omega^2} \cdot \frac{\int\limits_0^{\Omega} G(\omega)\, d\omega}{P_0\Omega} = 3\beta^2\gamma_{\text{вых } AM}, \tag{7.21}$$

где $\beta = \frac{\Delta\omega}{\Omega}$ — индекс частотной модуляции.

Таким образом, при произвольном спектре сигнала частотная модуляция по сравнению с амплитудной обеспечивает выигрыш в $3\beta^2$ раз. Этот выигрыш достигнут ценой расширения спектра модулированного сигнала.

Кодоимпульсная модуляция

При кодоимпульсной модуляции имеют место два вида помех:
шум квантования $\delta_к(t)$;
внешние помехи $\xi(t)$.

Оценим влияние каждой из указанных помех. Шум квантования представляет собой методическую погрешность, возникающую при квантовании сигналов по уровню. Как уже установлено (см. параграф 3.10), при условии, что шаг квантования Δ_x значительно меньше динамического диапазона сигнала, шум квантования распределен по закону равной вероятности в пределах шага квантования, причем дисперсия шума квантования

$$D(\delta_к) = \frac{\Delta_x^2}{12}.$$

Пусть величина сигнала распределена равновероятно в интервале от 0 до A. Тогда дисперсия сигнала

$$D(X) = \int\limits_0^A X^2 \omega(X)\, dX = \frac{A^2}{3}.$$

Отношение мощности сигнала к мощности шума квантования

$$\gamma_{КИМ} = \frac{D(X)}{D(\delta_к)} = 4\frac{A^2}{\Delta_x^2}. \tag{7.22}$$

Число уровней квантования

$$m = \frac{A}{\Delta_x}.$$

Если квантованные сигналы кодируются двоичным кодом, то

$$m = \frac{A}{\Delta_x} = 2^n, \qquad (7.23)$$

где n — количество разрядов в кодовой комбинации.

Тогда

$$\gamma_{КИМ} = 4m^2 = 2^{2n+2}. \qquad (7.24)$$

Если модулируемый сигнал имеет полосу частот $0 \div F_c$, то квантование сигнала по времени должно осуществляться с частотой $2F_c$. Пусть продолжительность кодовых комбинаций равна интервалу временного квантования сигнала $T_к = \frac{1}{2F_c}$. Тогда при условии, что длительность одного импульса τ в кодовой комбинации равна длительности паузы между импульсами, получим следующее выражение для ширины частотного спектра модулированного сигнала:

$$\Delta f_к \cong \frac{1}{\tau} = \frac{2n}{T_к} = 4F_c n,$$

откуда

$$n = 4\frac{\Delta f_к}{F_c}. \qquad (7.25)$$

Как видно из (7.25), коэффициент n выражает отношение спектра модулированного сигнала к спектру исходящего немодулированного сигнала, т. е. играет такую же роль, как и индекс модуляции β при частотной модуляции.

Таким образом, можно сделать заключение, что при кодоимпульсной модуляции отношение мощностей сигнала и помехи пропорционально $2^{2\beta}$.

Высокая помехоустойчивость кодоимпульсной модуляции достигнута как за счет расширения спектра модулированного сигнала, так и за счет увеличения времени передачи.

Незначительное увеличение спектра сигнала с КИМ (увеличение n) сопровождается резким увеличением помехоустойчивости данного вида модуляции. Однако, как видно из (7.23), рост n (при заданном A) может быть осуществлен за счет существенного уменьшения шага квантования Δ_x. Последнее же может, в свою очередь, привести к понижению помехоустойчивости КИМ за счет влияния внешних помех.

Действительно, пусть в канале действует аддитивная помеха $\xi(t)$. Действие помехи $\xi(t)$ приводит к ошибкам при квантовании сигнала по уровню, если уровень помехи превосходит половину шага квантования. Вероятность такой ошибки

$$p_{ош} = \int_{\frac{\Delta_x}{2}}^{\infty} w(\xi)\, d\xi + \int_{-\infty}^{-\frac{\Delta_x}{2}} w(\xi)\, d\xi = 2\int_{\frac{\Delta_x}{2}}^{\infty} w(\xi)\, d\xi.$$

При нормальном законе распределения помехи

$$p_{\text{ош}} = \frac{2}{\sqrt{2\pi\sigma_\xi^2}} \int_0^\infty \exp\left\{-\frac{\xi^2}{2\sigma_\xi^2}\right\} d\xi = 1 - \Phi\left(\frac{\Delta_x}{2\sqrt{2}\,\sigma_\xi}\right), \qquad (7.26)$$

где $\Phi\left(\frac{\Delta_x}{2\sqrt{2}\,\sigma_\xi}\right)$ — функция Лапласа; σ_ξ — среднеквадратическое значение помехи.

На рис. 7.2 показаны графики изменения функций Лапласа и вероятности ошибок в зависимости от отношения $\frac{\Delta_x}{2\sqrt{2}\,\sigma_\xi}$. Как

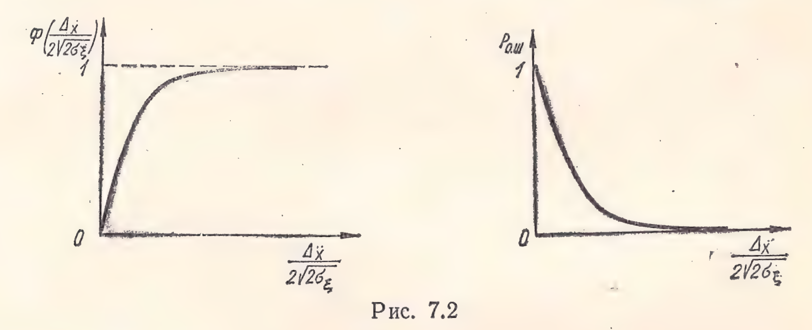


Рис. 7.2

видно, при увеличении шага квантования Δ_x происходит резкое уменьшение вероятностей ошибок за счет действия внешних помех.

Таким образом, выбор шага квантования Δ_x при КИМ должен осуществляться с учетом влияния обоих видов помех: внешних помех и шума квантования.

7.5. Помехоустойчивое кодирование. Основные принципы помехоустойчивого кодирования

Помехоустойчивые коды — одно из наиболее эффективных средств обеспечения высокой верности передачи дискретной информации. Создана специальная теория помехоустойчивого кодирования, быстро развивающаяся в последнее время.

Бурное развитие теории помехоустойчивого кодирования связано с внедрением автоматизированных систем, у которых обработка принимаемой информации осуществляется без участия человека. Использование для обработки информации электронных цифровых вычислительных машин предъявляет очень высокие требования к верности передачи информации.

Теорема Шеннона для дискретного канала с помехами утверждает, что вероятность ошибок за счет действия в канале помех может быть обеспечена сколь угодно малой путем выбора соответствующего способа кодирования сигналов. Из этой теоремы вытекает

весьма важный вывод о том, что наличие помех не накладывает принципиально ограничений на верность передачи.

Однако в теореме Шеннона не говорится о том, как нужно строить помехоустойчивые коды.

На этот вопрос отвечает теория помехоустойчивого кодирования.

Рассмотрим сущность помехоустойчивого кодирования, а также некоторые теоремы и определения, относящиеся к теории такого кодирования.

Под помехоустойчивыми или корректирующими кодами понимают коды, позволяющие обнаружить и устранить ошибки, происходящие при передаче из-за влияния помех.

Для выяснения идеи помехоустойчивого кодирования рассмотрим двоичный код, нашедший на практике наиболее широкое применение.

Напомним, что двоичный код — это код с основанием $m = 2$. Количество разрядов n в кодовой комбинации принято называть длиной или значностью кода. Каждый разряд может принимать значения 0 или 1. Количество единиц в кодовой комбинации называют весом кодовой комбинации и обозначают w.

Например, кодовая комбинация 100101100 характеризуется значностью $n = 9$ и весом $w = 4$.

Степень отличия любых двух кодовых комбинаций данного кода характеризуется так называемым расстоянием между кодами d. Оно выражается числом позиций или символов, в которых комбинации отличаются одна от другой. Кодовое расстояние есть минимальное расстояние между кодовыми комбинациями данного кода, оно определяется как вес суммы по модулю два кодовых комбинаций. Например, для определения расстояния между комбинациями 100101100 и 110110101 необходимо просуммировать их по модулю два

$$\begin{array}{r} \oplus\begin{array}{r}100101100\\110110101\end{array}\\ \hline 010011001\end{array}$$

Полученная в результате суммирования новая кодовая комбинация характеризуется весом $w = 4$. Следовательно, расстояние между исходными кодовыми комбинациями $d = 4$.

Ошибки, вследствие воздействия помех, проявляются в том, что в одном или нескольких разрядах кодовой комбинации нули переходят в единицы и, наоборот, единицы переходят в нули. В результате создается новая ложная кодовая комбинация.

Если ошибки происходят только в одном разряде кодовой комбинации, то такие ошибки называются однократными. При наличии ошибок в двух, трех и т. д. разрядах ошибки называются двукратными, трехкратными и т. д.

Для указания мест в кодовой комбинации, где имеются искажения символов, используется вектор ошибки $\bar{e}$. Вектор ошибки n-разрядного кода — это n-разрядная комбинация, единицы в которой

указывают положение искаженных символов кодовой комбинации. Например, если для пятиразрядного кода вектор ошибки имеет вид $\bar{e} = 01100$, то это значит, что имеют место ошибки в третьем и четвертом разрядах кодовой комбинации.

Вес вектора ошибки w_e характеризует кратность ошибки. Сумма по модулю для искаженной кодовой комбинации и вектора ошибки дает исходную неискаженную комбинацию.

Помехоустойчивость кодирования обеспечивается за счет введения избыточности в кодовые комбинации. Это значит, что из n символов кодовой комбинации для передачи информации используется $k < n$ символов. Следовательно, из общего числа $N_0 = 2^n$ возможных кодовых комбинаций для передачи информации используется только $N = 2^k$ комбинаций. В соответствии с этим все множества $N_0 = 2^n$ возможных кодовых комбинаций делятся на две группы. В первую группу входит множество $N = 2^k$ разрешенных комбинаций. Вторая группа включает в себя множество $(N_0 - N) = 2^n - 2^k$ запрещенных комбинаций.

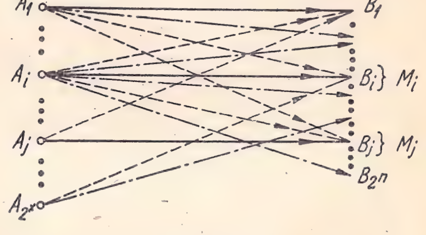


Рис. 7.3

Если на приемной стороне установлено, что принятая комбинация относится к группе разрешенных, то считается, что сигнал пришел без искажений. В противном случае делается вывод, что принятая комбинация искажена. Однако это справедливо лишь для таких помех, когда исключена возможность перехода одних разрешенных комбинаций в другие.

В общем случае каждая из N разрешенных комбинаций может трансформироваться в любую из N_0 возможных комбинаций, т. е. всего имеется $N \cdot N_0$ возможных случаев передачи (рис. 7.3), из них N случаев безошибочной передачи (на рис. 7.3 обозначены жирными линиями), $N(N-1)$ случаев перехода в другие разрешенные комбинации (на рис. 7.3 обозначены пунктирными линиями) и $N(N_0 - N)$ случаев перехода в запрещенные комбинации (на рис. 7.3 обозначены штрих пунктирными линиями).

Таким образом, не все искажения могут быть обнаружены. Доля обнаруживаемых ошибочных комбинаций составляет

$$\frac{N(N_0 - N)}{N \cdot N_0} = 1 - \frac{N}{N_0}. \tag{7.27}$$

Для использования данного кода в качестве исправляющего множество запрещенных кодовых комбинаций разбивается на N непересекающихся подмножеств $M_к$. Каждое из подмножеств $M_к$ ставится в соответствие одной из разрешенных комбинаций.

Если принятая запрещенная комбинация принадлежит подмножеству M_i, то считается, что передана комбинация A_i (рис. 7.3).

Ошибка будет исправлена в тех случаях, когда полученная комбинация действительно образовалась из комбинации A_i. Таким образом, ошибка исправляется в $(N_0 - N)$ случаях, равных количеству запрещенных комбинаций. Доля исправляемых ошибочных комбинаций от общего числа обнаруживаемых ошибочных комбинаций составляет

$$\frac{N_0 - N}{N(N_0 - N)} = \frac{1}{N}. \tag{7.28}$$

Способ разбиения на подмножества зависит от того, какие ошибки должны исправляться данным кодом.

7.6. Связь исправляющей способности кода с кодовым расстоянием

Для оценки степени различия между двумя произвольными комбинациями данного кода используется, как уже отмечалось, характеристика, получившая название расстояния между кодовыми комбинациями. Наименьшее расстояние между разрешенными кодовыми комбинациями называют кодовым расстоянием и обозначают $d_{мин}$. Это очень важная характеристика кода, ибо именно она характеризует его корректирующие способности.

Рассмотрим это на конкретных примерах.

Пусть необходимо построить код, обнаруживающий все ошибки кратностью t и ниже.

Построить такой код — это значит из множества N_0 возможных комбинаций выбрать N разрешенных комбинаций так, чтобы любая из них в сумме по модулю два с любым вектором ошибок с весом $w_e \leqslant t$ не дала бы в результате никакой другой разрешенной комбинации. Для этого необходимо, чтобы кодовое расстояние удовлетворяло условию

$$d_{мин} \geqslant t + 1. \tag{7.29}$$

В качестве примера рассмотрим код со значностью $n = 3$. Все возможные комбинации такого кода представлены в табл. 14. Матрица расстояний между кодовыми комбинациями имеет вид (табл. 15).

Таблица 14

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
000	001	010	011	100	101	110	111

Для того чтобы код обеспечивал обнаружение однократных ошибок, необходимо из всего множества $N_0 = 8$ возможных комбинаций выбрать в качестве разрешенных такие, расстояние между

Таблица 15

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
A_1	0	1	1	2	1	2	2	3
A_2		0	2	1	2	1	3	2
A_3			0	1	2	3	1	2
A_4				0	3	2	2	1
A_5					0	1	1	2
A_6						0	2	1
A_7							0	1
A_8								0

которыми было бы не менее $d = 2$. Как видно из матрицы расстояний, в качестве разрешенных комбинаций в этом случае можно выбрать следующие:

$$A_1 = 000; \quad A_4 = 011; \quad A_6 = 101; \quad A_7 = 110.$$

Для обнаружения двукратных ошибок кодовое расстояние должно быть равно $d_{\text{мин}} = 3$. При этом в качестве разрешенных комбинаций можно выбрать

$$A_1 = 000; \quad A_2 = 111.$$

Очевидна справедливость условия

$$d \leqslant n. \qquad (7.30)$$

Следовательно, в данном случае двукратные ошибки не могут быть обнаружены, так как в результате таких ошибок одни разрешенные комбинации переходят в другие со значностью $n = 3$. Пусть теперь необходимо построить код, обеспечивающий устранение однократных ошибок. Выбираем в качестве первой разрешенной комбинации $A_1 = 000$. При наличии однократных ошибок комбинация A_1 может перейти в одну из следующих запрещенных комбинаций: $A_2 = 001$, $A_3 = 010$ и $A_5 = 100$. Комбинации A_2, A_3 и A_5 можно принять в качестве подмножества запрещенных комбинаций

комбинации A_1. Последнее означает, что в случае приема одной из комбинаций этого подмножества выносится решение, что передана комбинация A_1.

Пусть в качестве второй разрешенной комбинации выбирается комбинация, отстоящая от первой на расстоянии $d = 2$, например $A_4 = 011$. Ей должно соответствовать подмножество запрещенных комбинаций $A_3 = 010$; $A_2 = 001$ и $A_8 = 111$. Однако получилось пересечение подмножеств. При приеме запрещенных сигналов A_2 или A_3 нельзя однозначно установить, какой был передан сигнал — A_1 или A_4.

Если же в качестве второй разрешенной комбинации выбрать комбинацию, отстоящую от A_1 на $d = 3$, т. е. комбинацию $A_8 = 111$, которой соответствует подмножество запрещенных комбинаций $A_4 = 011$, $A_6 = 101$ и $A_7 = 110$, то в этом случае подмножества запрещенных комбинаций не пересекаются. Следовательно, при $d = 3$ обеспечивается устранение всех однократных ошибок.

В общем случае для устранения ошибок кратности σ кодовое расстояние должно удовлетворять условию

$$d_{\text{мин}} \geqslant 2\sigma + 1. \tag{7.31}$$

Аналогично рассуждая, можно установить, что для исправления всех ошибок кратности не более σ и одновременного обнаружения всех ошибок кратности не более t (при $t \geqslant \sigma$) кодовое расстояние должно удовлетворять условию

$$d_{\text{мин}} \geqslant t + \sigma + 1. \tag{7.32}$$

При этом нужно иметь в виду, что если обнаруженная кодом ошибка имеет кратность $t > \sigma$, то такая ошибка исправлена быть не может, т. е. в данном случае код только обнаруживает ошибку.

7.7. Построение кодов с заданной исправляющей способностью

До сих пор при рассмотрении корректирующих кодов мы предполагали заданной его значность n. Повышение корректирующей способности кода достигалось при сохранении n за счет уменьшения множества N разрешенных комбинаций (или уменьшения количества k информационных символов). Обычно же на практике коды строятся в обратном порядке: вначале выбирается количество информационных символов k, исходя из объема алфавита источника, а затем обеспечивается необходимая корректирующая способность кода за счет добавления избыточных символов.

Пусть известен объем алфавита источника N. Необходимое количество информационных символов определяется из выражения

$$k = \log_2 N.$$

Пусть также известно полное число векторов ошибок E, которое необходимо исправить.

Задача состоит в том, чтобы при заданных N и E определить значность кода n, обладающего требуемыми корректирующими возможностями.

Полное число ошибочных комбинаций, подлежащих исправлению, равно $E \cdot 2^k = E \cdot N$. Так как количество ошибочных комбинаций равно $N_0 - N$, то код обеспечивает исправление не более $N_0 - N$ комбинаций. Следовательно, необходимое условие для возможности исправления ошибок можно записать в виде

$$NE \leqslant N_0 - N,$$

откуда получим

$$N_0 \geqslant (1 + E) N, \tag{7.33}$$

или

$$N \leqslant \frac{2^n}{1 + E}. \tag{7.34}$$

Формула (7.34) выражает условие для выбора значности кода n.

Рассмотрим частные случаи. Если имеются ошибки разной кратности, то прежде всего необходимо обеспечить устранение однократных ошибок, вероятность появления которых наибольшая. Возможное количество векторов однократных ошибок

$$E = C_n^1 = n.$$

В этом случае зависимость (7.34) примет вид

$$2^k = N \leqslant \frac{2^n}{1 + n}. \tag{7.35}$$

При построении кода целесообразно пользоваться табл. 16.

Таблица 16

n	2	3	4	5	6	7	8	9
$\frac{2^n}{1+n}$	1,33	2	3,2	5,33	9,2	16	28,4	51,2

Нужно при этом иметь в виду, что код должен также удовлетворять условию

$$d_{\text{мин}} \geqslant 3. \tag{7.36}$$

Если необходимо обеспечить устранение всех ошибок кратности от 1 до l, то нужно учесть, что

число возможных однократных ошибок $E_1 = C_n^1$;
число возможных двукратных ошибок $E_2 = C_n^2$;
. . . .
число возможных l-кратных ошибок $E_l = C_n^l$.

Общее число ошибок $E = \sum_{i=1}^{l} C_n^i$.

При этом зависимость (7.34) с определенным приближением примет вид

$$N \leqslant \frac{2^n}{1 + \sum_{i=1}^{l} C_n^i}. \qquad (7.37)$$

7.8. Геометрическая модель комбинаций двоичного кода

Комбинации n-разрядного двоичного кода можно рассматривать как вершины n-мерного единичного куба, т. е. куба с длиной ребра, равной 1.

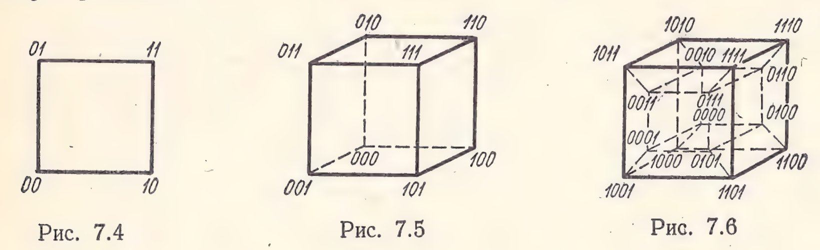


Рис. 7.4 Рис. 7.5 Рис. 7.6

При $n = 2$ точки, отображающие кодовые комбинации, располагаются в вершинах единичного квадрата (рис. 7.4); при $n = 3$ в вершинах единичного куба (рис. 7.5), при $n = 4$ в вершинах четырехмерного куба (рис. 7.6) и т. д.

Искажения кодовых комбинаций интерпретируются на геометрической модели перемещением отображающих точек из одних вершин куба в другие. Кратность ошибки интерпретируется количеством ребер куба, на которое отстоит точка, отображающая искаженную комбинацию, от точки, отображающей исходную неискаженную комбинацию.

Вес кодовой комбинации интерпретируется на геометрической модели числом ненулевых координат точки кода.

Расстоянию между кодовыми комбинациями соответствует на геометрической модели число ребер куба, на которое отстоят друг от друга соответствующие отображающие точки.

Для исправления ошибок к подмножеству каждой разрешенной комбинации относят все вершины куба, лежащие в сфере радиусом $\frac{d-1}{2}$ и центром в вершине, соответствующей данной разрешенной комбинации.

В случае независимости ошибок точки, соответствующие разрешенным кодовым комбинациям, должны располагаться в пространстве равномерно.

7.9. Показатели количества корректирующих кодов

Одной из основных характеристик корректирующего кода является его способность обеспечить правильный прием кодовых комбинаций при наличии искажений под воздействием помех, т. е. помехоустойчивость кода. Помехоустойчивость кодов количественно оценивают величиной

$$S = \lg \frac{1}{p_{н.\,о}}, \qquad (7.38)$$

где $p_{н.\,о}$ — вероятность неправильного приема кодовой комбинации.

Часто для оценки помехоустойчивости кодов пользуются понятием коэффициента обнаружения

$$K_{обн} = \frac{L}{M}, \qquad (7.39)$$

где M — наиболее вероятное общее количество искаженных комбинаций из числа N переданных комбинаций; L — наиболее вероятное количество искаженных комбинаций, ошибки в которых обнаруживаются.

При передаче достаточно большого числа кодовых комбинаций N можно считать, что общее количество искаженных комбинаций M и количество искаженных комбинаций L, ошибки в которых обнаруживаются, соответственно

$$M \cong Np_к;$$
$$L \cong Np_{о.\,о},$$

где $p_к$ — вероятность искажений кодовой комбинации; $p_{о.\,о}$ — вероятность появления обнаруживаемых искажений комбинаций.

Тогда

$$K_{обн} = \frac{p_{о.\,о}}{p_к}. \qquad (7.40)$$

Важной характеристикой корректирующего кода является его избыточность, указывающая степень удлинения кодовой комбинации для достижения определенной корректирующей способности.

Для оценки избыточности корректирующего кода пользуются понятием коэффициента избыточности

$$r = \frac{\rho}{n} = \frac{n-k}{n}, \qquad (7.41)$$

где $\rho = n - k$ — количество избыточных позиций кодовой комбинации, используемых для обеспечения корректирующих способностей кода.

7.10. Классификация помехоустойчивых кодов

К настоящему времени создано довольно большое количество корректирующих кодов. Классификация известных корректирующих кодов может быть осуществлена следующим образом (рис. 7.7).

Все корректирующие коды можно разделить на два основных класса: непрерывные (рекурентные, или цепные) и блочные.

В непрерывных кодах процесс кодирования и декодирования носит непрерывный характер. Каждый избыточный (проверочный) символ формируется по двум или нескольким информационным символам. Проверочные символы размещаются в определенном порядке между информационными символами исходной последовательности. Этот класс кодов появился совсем недавно и не получил пока широкого развития.

В блочных кодах каждому сообщению (или элементу сообщения) сопоставляется кодовая комбинация (блик) из n символов. Блоки кодируются и декодируются отдельно друг от друга.

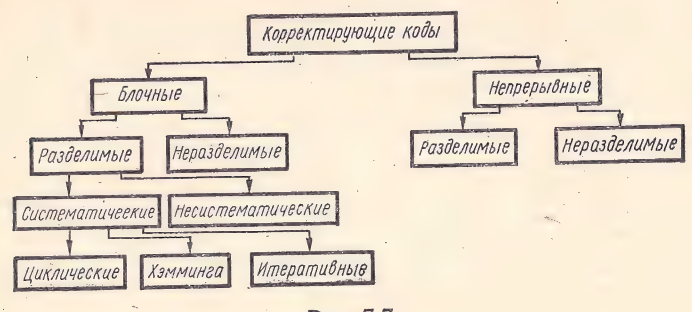


Рис. 7.7

Блочные коды могут быть равномерными, когда n остается постоянным для всех комбинаций, или неравномерными, когда n непостоянно.

Неравномерные корректирующие коды не получили практического применения из-за сложности их технической реализации.

Следует заметить, что рассмотренные ранее принципы помехоустойчивого кодирования, а также геометрическая модель комбинаций кода относятся к блочным равномерным кодам.

Как непрерывные, так и блочные коды в зависимости от методов внесения избыточности подразделяются на разделимые и неразделимые. В разделимых кодах четко разграничена роль отдельных символов. Одни символы являются информационными, другие являются проверочными и служат для обнаружения и исправления ошибок. Информационные и проверочные символы занимают во всех кодовых комбинациях одни и те же позиции. Разделимые блочные коды называются обычно n, k-кодами, где n — значность кода, k — число информационных символов.

Неразделимые коды не имеют четкого разделения кодовой комбинации на информационные и проверочные символы. Этот класс кодов пока немногочислен.

Разделимые блочные коды делятся, в свою очередь, на несистематические и систематические. В несистематических кодах про-

верочные символы представляют суммы подблоков длиной l, на которые разделена последовательность информационных символов. Такой код способен обнаружить серийные ошибки с длиной серии, не превосходящей l.

Самый большой класс разделимых блочных кодов составляют систематические, или линейные коды, у которых проверочные символы определяются в результате проведения линейных операций над определенными информационными символами. Для случая двоичных кодов каждый проверочный символ выбирается таким, чтобы его сумма по mod 2 с определенными информационными символами стала равной нулю. Декодирование сводится к проверке на четность определенных групп символов. В связи с этим такие коды получили другое название — коды с проверкой на четность. В результате проверок на четность определенных групп символов кодовой комбинации дается информация о наличии ошибок, а в случае необходимости — о позиции символов, где имеются ошибки.

Рассмотрим некоторые типы систематических кодов.

7.11. Код с четным числом единиц

Код содержит лишь один избыточный символ. Выбирается избыточный символ таким, чтобы его сумма по mod 2 со всеми информационными символами равнялась нулю. Благодаря такому способу выбора избыточного символа кодовая комбинация содержит четное число единиц.

Избыточность кода равна

$$r = \frac{p}{n} = \frac{1}{n},$$

где n — значность кода; p — число проверочных символов.

Признаком искажения кодовой комбинации является нечетность единиц в комбинации. Код позволяет обнаруживать однократные ошибки и все ошибки нечетной кратности, так как только в этих случаях количество единиц в комбинации станет нечетным.

7.12. Код с удвоением элементов и инверсный код

Код с удвоением элементов характеризуется введением дополнительных символов для каждого символа информационной части комбинации, причем единица дополняется нулем и преобразуется в 10, а нуль дополняется единицей и преобразуется в 01. Тогда исходная, например, комбинация 10101 будет представлена в виде 1001100110. Показателем искажения кода будет появление в «парных» элементах сочетаний вида 00 или 11.

Код позволяет обнаруживать все ошибки, за исключением случаев, когда имеют место двукратные ошибки в «парных» элементах.

Избыточность кода не зависит от числа элементов кодовой комбинации

$$r = 0{,}5.$$

Естественно, что помехоустойчивость этого кода выше, чем кода с четным числом единиц.

В основу построения инверсного кода положен метод повторения исходной кодовой комбинации. Причем в тех случаях, когда исходная комбинация содержит четное число единиц, вторая комбинация в точности воспроизводит исходную, если же исходная комбинация содержит нечетное число единиц, то повторение происходит в инвертированном виде. Например, комбинации 01010 и 01110 инверсным кодом представляются соответственно как 0101001010 и 0111010001.

Проверка кодовой комбинации производится с такой последовательностью. Сначала суммируются единицы, содержащиеся в основной комбинации. Если их число окажется четным, то элементы дополнительной комбинации принимаются в неизменном виде. После этого обе комбинации сравниваются поэлементно (первый элемент с первым, второй со вторым и т. д.) и при обнаружении хотя бы одного несовпадения принятая комбинация бракуется.

Если же количество единиц основной комбинации нечетное, элементы второй комбинации принимаются в инвертированном виде. Затем, как и в предыдущем случае, основная и дополнительная комбинации сравниваются поэлементно.

Такое построение кода позволяет обнаружить практически все ошибки в комбинации. Ошибки не будут обнаружены лишь тогда, когда одновременно исказятся два (четыре и т. д.) элемента в исходной комбинации. Очевидно, что вероятность таких ошибок чрезвычайно мала.

7.13. Коды Хемминга

Известно несколько разновидностей кода Хемминга, характеризуемых различной корректирующей способностью. Но в основу построения всех их положен один и тот же метод. Ограничимся рассмотрением лишь принципов построения кодов, исправляющих одиночные ошибки.

Код Хемминга, как и любой (n, k) код, содержит k информационных и $\rho = n - k$ избыточных символов. Избыточная часть кода строится таким образом, чтобы при декодировании можно было бы установить не только факт наличия ошибок в принятой комбинации, но и указать номер позиции, в которой произошла ошибка. Это достигается за счет многократной проверки принятой комбинации на четность. Количество проверок равно количеству избыточных символов ρ. Каждой проверкой должны охватываться часть информационных символов и один из избыточных символов. При каждой проверке получают двоичный контрольный символ. Если результат проверки дает четное число, то контрольному символу присваивается значение 0, если нечетное число — 1. В результате всех проверок получается ρ-разрядное двоичное число, указывающее номер искаженного символа. Для исправления ошибки достаточно лишь изменить значение данного символа на обратное.

Необходимое количество проверочных символов ρ (или значность кода n) определяется из соотношения (7.35). Значения проверочных символов и номера их позиций по методике Хемминга устанавливаются одновременно с выбором контролируемых групп кодовой комбинации. При этом нужно исходить из следующего.

В результате первой проверки получается цифра младшего разряда контрольного числа, указывающего номер искаженного символа. Если результат первой проверки даст 1, то один из символов проверенной группы искажен.

Для выяснения вопроса, какой из символов при этом может быть искажен, рассмотрим табл. 17, в которой представлен для иллюстрации натуральный ряд четырехзначных контрольных чисел в двоичной системе счисления. Как видно из таблицы, если младший разряд контрольного числа содержит единицу, то искажение должно быть в одной из нечетных позиций кодовой комбинации. Следовательно, первой проверкой должны быть охвачены символы с нечетными номерами: 1, 3, 5, 7, 9 и т. д. (Порядковые номера символов в кодовой комбинации читаются слева направо).

Таблица 17

№ п. п.	Символы разрядов контрольного числа			
	4	3	2	1
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0

Если результат второй проверки даст 1, то получим 1 во втором разряде контрольного числа. Следовательно, второй проверкой должны быть охвачены символы с номерами, содержащими в двоичной записи единицы во втором разряде: 2, 3, 6, 7, 10.

Аналогично при третьей проверке должны проверяться символы, номера которых в двоичной записи содержат единицы в третьем разряде: 4, 5, 6, 7, 12 и т. д.

Такие рассуждения позволяют образовать табл. 18 проведения проверок.

Если символы проверяемой кодовой комбинации обозначить через a_i, то проверочные операции S_i можно выразить следующим образом:

$$
\begin{aligned}
S_1 &= a_1 \oplus a_3 \oplus a_5 \oplus a_7 \oplus a_9 \oplus \cdots \\
S_2 &= a_2 \oplus a_3 \oplus a_6 \oplus a_7 \oplus a_{10} \oplus \cdots \\
S_3 &= a_4 \oplus a_5 \oplus a_6 \oplus a_7 \oplus a_{12} \oplus \cdots \\
S_4 &= a_8 \oplus a_9 \oplus a_{10} \oplus a_{11} \oplus a_{12} \oplus \cdots \\
&\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots
\end{aligned}
\qquad (7.42)
$$

С целью упрощения операций кодирования и декодирования целесообразно выбирать такое размещение проверочных символов в кодовой комбинации, при котором каждый из них включается

в минимальное число проверяемых групп символов. В связи с этим удобно размещать контрольные символы на позициях, номера которых встречаются только в одной из проверяемых групп: 1, 2, 4, 8, ... (см. табл. 18). Следовательно, в кодовой комбинации символы a_1, a_2, a_4, a_8 ... должны быть проверочными и символы a_3, a_5, a_6, a_7, a_9 ... — информационными.

Так как значения информационных символов проставляются заранее, то значения проверочных символов должны быть такими, чтобы сумма единиц в каждой проверочной группе являлась четным числом.

Представим в качестве примера простую двоичную комбинацию 10011 кодом Хэмминга. При числе информационных символов $k = 5$ в соответствии с (7.35) и табл. 17 позиционность кода Хэмминга должна быть $n = 9$. Так как информационными должны быть третий, пятый, шестой, седьмой, девятый символы, то для рассматриваемого кода $a_3 = 1$, $a_5 = 0$, $a_6 = 0$, $a_7 = 1$, $a_9 = 1$. Из условия обеспечения четности сумм (7.42) получим следующие значения проверочных символов: $a_1 = 1$, $a_2 = 0$, $a_4 = 1$; $a_8 = 1$. Следовательно, простому пятиэлементному коду 10011 соответствует девятиэлементный код Хемминга 101100111.

Таблица 18

№ проверки	Номера проверяемых позиций	Номера позиций контрольных символов
1	1, 3, 5, 7, 9, 11, 13	1
2	2, 3, 6, 7, 10	2
3	4, 5, 6, 7, 12	4
4	8, 9, 10, 11, 12 . . .	8
.	. . . .	.

Пусть теперь при передаче произошла ошибка в пятом символе, т. е. код принял вид 101110111. Тогда в результате первой проверки получим 1, второй — 0, третьей — 1 и четвертой — 0. Таким образом, в результате проверок получено контрольное двоичное число 0101, указывающее на искажение пятого символа.

7.14. Циклические коды

Из всех известных корректирующих кодов циклические коды являются наиболее простыми и эффективными. Эти коды могут быть использованы как для обнаружения и исправления независимых ошибок, так и, в особенности, для обнаружения и исправления серийных ошибок. Схемы кодирующих и декодирующих устройств для этих кодов чрезвычайно просты и представляют собой обычные регистры сдвига. Основное их свойство состоит в том, что каждая кодовая комбинация может быть получена путем циклической перестановки символов комбинации, принадлежащей к этому же коду. Это значит, что если кодовый вектор $v = (a_0, a_1, a_2, \ldots, a_{n-1})$ принадлежит к циклическому коду V, то вектор v', получаемый из v циклической перестановкой составляющих, т. е. $v'' = (a_{n-1}, a_0, a_1, \ldots, a_{n-2})$, также принадлежит коду V.

Рассмотрение циклических кодов более удобно производить, представляя комбинацию двоичного кода не в виде последовательностей нулей и единиц, а в виде полинома некоторой степени, а именно

$$G(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0, \qquad (7.43)$$

где x — фиктивная переменная; a_i — цифры данной системы счисления (в двоичной системе 0 и 1).

Так, например, двоичное семиразрядное число 1010101 может быть записано в виде полинома от переменной x

$$1010101 = G(x) = 1 \cdot x^6 + 0 \cdot x^5 + 1 \cdot x^4 + 0 \cdot x^3 + 1 \cdot x^2 + \\ + 0 \cdot x + 1 \cdot x^0 = x^6 + x^4 + x^2 + 1. \qquad (7.44)$$

Представление кодовых комбинаций в форме (7.44) позволяет свести действия над комбинациями к действиям над многочленами. При этом сложение двоичных многочленов сводится к сложению по модулю 2 коэффициентов при равных степенях переменной x; умножение производится по обычному правилу перемножения степенных функций, однако полученные при этом коэффициенты при данной степени складываются по модулю 2; деление осуществляется по правилам деления степенных функций, при этом операции вычитания заменяются операциями суммирования по модулю 2.

Представление комбинаций в форме (7.43) и (7.44) удобно еще и тем, что упомянутая ранее циклическая перестановка есть результат простого умножения данного полинома на x. Действительно, если одна из кодовых комбинаций выражается полиномом $v(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$, то новая комбинация за счет циклического сдвига будет $x \cdot v(x) = a_0x + a_1x^2 + a_2x^3 + \cdots + a_{n-2}x^{n-1} + a_{n-1}x^n$. В последнем члене необходимо заменить x^n на единицу. Следовательно, новая комбинация будет $v'(x) = a_{n-1} + a_0x + a_1x^2 + \cdots + a_{n-2}x^{n-1}$.

Используя представление двоичных кодов в виде полиномов, можно дать следующее определение циклическим кодам.

Циклический код — это такой (n, k) код, который образуется путем умножения простого k-значного кода, выраженного в виде полинома $Q(x)$ степени $k - 1$, на некоторый образующий полином $P(x)$ степени $(n - k)$.

В результате умножения всех кодовых комбинаций простого k-значного кода на образующий полином $P(x)$ число разрешенных комбинаций не изменяется и остается равным 2^k, а общее число запрещенных кодовых комбинаций будет равно $2^n - 2^k$.

Разрешенные комбинации циклического кода образуют некоторое множество комбинаций, отличающееся тем, что отображающие их полиномы делятся без остатка на образующий полином $P(x)$. При делении полиномов запрещенных комбинаций на образующий полином $P(x)$ обязательно появится остаток.

Это свойство циклического кода используется для исправления или для обнаружения ошибок. Действительно, если под воздействием

помех разрешенная кодовая комбинация трансформируется в запрещенную, то ошибка может быть обнаружена по остатку при делении комбинации на образующий полином $P(x)$.

Процедура построения циклического кода следующая [40]. Кодовая комбинация простого k-значного кода $G(x)$ умножается на одночлен x^{n-k}, а затем делится на образующий полином $P(x)$, степень которого равна $n-k$. В результате умножения комбинации $G(x)$ на x^{n-k} степень каждого одночлена, входящего в $G(x)$, повысится на $n-k$. При делении произведения $x^{n-k}\cdot G(x)$ на образующий полином $P(x)$ получится частное $Q(x)$ такой же степени, как и $G(x)$.

Результат умножения и деления можно представить в следующем виде:

$$\frac{x^{n-k}G(x)}{P(x)} = Q(x) + \frac{R(x)}{P(x)}, \qquad (7.45)$$

где $R(x)$ — остаток от деления $x^{n-k}G(x)$ на $P(x)$.

Так как частное $Q(x)$ имеет такую же степень, как и кодовая комбинация $G(x)$ простого кода, то $Q(x)$ является кодовой комбинацией того же простого k-значного кода.

Умножая обе части равенства (7.45) на $P(x)$ и произведя некоторые перестановки, получим

$$F(x) = Q(x)P(x) = x^{n-k}G(x) + R(x). \qquad (7.46)$$

В правой части (7.46) знак минус перед $R(x)$ заменен знаком плюс, так как вычитание по модулю 2 сводится к сложению.

Таким образом, кодовая комбинация циклического n-значного кода может быть получена двумя способами:

1) путем умножения кодовой комбинации $G(x)$ простого кода на одночлен x^{n-k} и добавления к этому произведению остатка $R(x)$, полученного в результате деления произведения $x^{n-k}G(x)$ на образующий полином $P(x)$;

2) путем умножения кодовой комбинации $Q(x)$ простого k-значного кода на образующий полином $P(x)$.

При первом способе кодирования первые k символов полученной кодовой комбинации совпадают с соответствующими символами исходного простого кода.

При втором способе в полученном коде информационные символы не всегда совпадают с символами исходного простого кода.

Такой способ легко реализуем, но вследствие того, что в полученных кодовых комбинациях не содержатся информационные символы в явном виде, усложняется процесс дешифрации.

После исправления ошибок такие комбинации для выделения информационных символов приходится делить на образующий многочлен.

Боуз и Чоудхури показали [40, 41], что для любых целых положительных чисел z и σ существует циклический код значности

$$n = 2^z - 1 \qquad (7.47)$$

и кодовое расстояние

$$d_{мин} \geqslant 2\sigma + 1. \qquad (7.48)$$

При этом число проверочных символов $\rho = n - k$ не превышает величины $z \cdot \sigma$, т. е.

$$\rho \leqslant z \cdot \sigma. \qquad (7.49)$$

Такой код гарантийно исправляет ошибки кратности σ и менее, или обнаруживает ошибки кратности 2σ и менее. Кроме того, код обнаруживает все пакеты ошибок, длина которых равна или меньше ρ.

Соотношения (7.47), (7.48) и (7.49) могут быть использованы для выбора образующего полинома, который нужно производить с учетом того, что степень полинома должна быть равна числу проверочных символов $\rho = n - k$. Кроме того, полином $P(x)$ должен входить в качестве сомножителя в разложение двучлена

$$(x^n + 1) = x^{2^z - 1} + 1. \qquad (7.50)$$

Двучлены типа (7.50) обладают тем свойством, что они являются общими кратными для всех без исключения неприводимых (т. е. не делящихся ни на какой другой многочлен) полиномов степени z и разлагаются на множители из всех неприводимых полиномов, степени z_i которых делят без остатка число z.

Рассмотрим в качестве примера метод построения циклического кода, содержащего $k = 4$ информационных символа и обеспечивающего устранение однократных ошибок или обнаружение двукратных ошибок.

В соответствии с (7.35) и табл. 15 определяем значность кода $n = 7$ и количество проверочных символов $\rho = 3$.

Для построения циклического кода необходимо выбрать образующий полином $P(x)$ степени $\rho = 3$, который, как указывалось ранее, должен входить в качестве сомножителя в разложение двучлена $x^n + 1 = x^{2^z + 1} + 1$. Так как $n = 7$, то этот двучлен имеет вид $x^7 + 1$. Составляющие его сомножители должны быть неприводимыми полиномами, степени которых являются делителями числа $z = 3$. К числам, на которые $z = 3$ делится без остатка, относятся 1 и 3. Следовательно, сомножителями двучлена $x^7 + 1$ должны быть неприводимые полиномы первой и третьей степеней.

Пользуясь таблицами неприводимых полиномов, имеющимися в [42] и [43], получим

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Один из сомножителей третьей степени, например $x^3 + x^2 + 1$, можно взять в качестве образующего полинома.

Для построения циклического кода остается умножить каждую простую четырехсимвольную комбинацию $Q(x)$ на образующий полином. Возьмем в качестве примера простую четырехсимвольную

комбинацию $Q(x)=0011$. Операция умножения этой комбинации на образующий полином $P(x)=1101$ запишется следующим образом:

$$\begin{array}{r} \times \begin{array}{r} 0011 \\ 1101 \end{array} \\ \hline 0011 \\ 0011 \\ 0011 \\ \hline 0011111 \end{array}$$

Таким образом, простая четырехсимвольная комбинация $Q(x) = 0011$ представляется семисимвольным циклическим кодом $F(x) = Q(x) \cdot P(x) = 0011111$.

Проиллюстрируем на примере процесс обнаружения двукратных ошибок. Пусть при передаче пятой комбинации 0101 имели место двукратные ошибки, представляемые вектором $e = 0100100$. Следовательно, будет принята комбинация

$$C_T^*(x) = C_T(x) \oplus \bar{e} = 0011101.$$

Произведя декодирование принятой комбинации путем деления ее на образующий полином $P(x) = 1101$, получим остаток $R(x) = 111$, степень которого меньше степени образующего полинома. Отличие $R(x)$ от нуля свидетельствует о наличии ошибок в принятой комбинации.

Более подробные сведения о циклических кодах даны в [28], [40], [41], [42] и [43].

7.15. Системы с обратной связью

Одним из эффективных способов повышения помехоустойчивости передачи является использование обратного канала, по которому информация может передаваться от приемника к передатчику. Системы, в которых применяется передача по обратному каналу, носят общее название систем с обратной связью.

Существует большое число способов построения обратного канала. Основные варианты схем обратной связи показаны на рис. 7.8 [37]. В варианте I обратная связь охватывает только линию связи. В вариантах II и III обратная связь подключена после решающего устройства.

Вариант III отличается тем, что обратная связь охватывает всю систему.

В зависимости от способа использования обратного канала системы с обратной связью могут быть разделены на два основных типа:

системы с информационной обратной связью (системы со сравнением);

системы с решающей обратной связью (системы с переспросом),

В системах с информационной обратной связью по обратному каналу передаются все принятые сигналы. На передающей стороне эти сигналы сравниваются с переданными. В случае наличия расхождений осуществляется повторная передача сигнала либо передача данных о необходимых исправлениях.

В системах с решающей обратной связью приемник сам проверяет правильность принимаемого сигнала. При обнаружении искажений приемник посылает по обратному каналу сигнал переспроса на повторение передачи искаженного сигнала.

Таким образом, в системах с информационной обратной связью решение о необходимости повторения передачи принимается на передающей стороне, а в системах с решающей обратной связью — на приемной стороне.

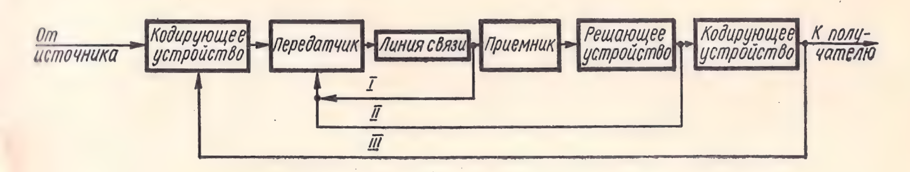


Рис. 7.8

Обратные каналы в системах с информационной обратной связью и в системах с решающей обратной связью используются не в одинаковой степени. Обратный канал в системе с решающей обратной связью загружен обычно значительно меньше, чем в системах с информационной обратной связью, ибо в системах с решающей обратной связью сигнал по обратному каналу поступает лишь в случае обнаружения искажений в переданном сигнале, в то время как в системах с информационной обратной связью каждый принятый сигнал пересылается по обратному каналу на передающую сторону.

Структура сигналов в системах с решающей обратной связью должна быть таковой, чтобы ошибка при передаче могла быть обнаружена на приемной стороне. Это может быть осуществлено лишь при использовании кодов, позволяющих обнаруживать искажения. Исследования показали [40], [44], что системы с решающей обратной связью, использующие коды с обнаружением ошибок, эффективнее систем без обратной связи, использующих коды с исправлением ошибок. Это объясняется тем, что в системах с решающей обратной связью повторение передачи осуществляется лишь при обнаружении ошибок в переданном сигнале, в то время, как в системах с автоматической коррекцией ошибок на приемной стороне независимо от наличия искажений в сигналах постоянно вводится избыточность, требуемая для исправления ошибок.

Более детально с этими системами можно ознакомиться в [37], [40], [44], [45] и др.

7.16. Примеры

Пример 7.1. Передана кодовая комбинация 1100. Известно, что вес вектора ошибки $w_e = 2$. Определить: 1) возможные варианты искаженных комбинаций, 2) кодовое расстояние, небходимое для обнаружения и устранения всех ошибок.

Решение. 1) искаженные комбинации получаются суммированием по модулю 2 исходной комбинации и вектора ошибки.

При весе $w_e = 2$ возможны следующие варианты вектора ошибки (при длине кода $n = 4$): $\bar{e} = 0011$, 0101; 1001; 0110; 1010; 1100.

При этом будут иметь место следующие варианты искаженных комбинаций: 1111; 1001; 0101; 1010; 0110; 0000;

2) при $w_e = 2$ кратность ошибок равна 2.

Для обнаружения таких ошибок необходимое кодовое расстояние

$$d_{\text{мин}} = t + 1 = 2 + 1 = 3.$$

Для устранения таких ошибок необходимо

$$d_{\text{мин}} = 2\sigma + 1 = 2 \cdot 2 + 1 = 5.$$

Для кода с $n = 4$ обеспечить кодовое расстояние $d_{\text{мин}} = 5$ невозможно. Следовательно, в нашем коде в принципе невозможно обеспечить устранение двойных ошибок.

Пример 7.2. Принята кодовая комбинация 0101. Известно, что вектор ошибки равен $\bar{e} = 0011$. Определить исходную неискаженную комбинацию.

Решение. Исходная комбинация может быть получена суммированием по модулю 2 искаженной комбинации и вектора ошибки

$$\begin{array}{r} \oplus\, 0101 \\ 0011 \\ \hline 0110 \end{array}$$

Таблица 19

	00000	01110	10101	11011
00000	0	3	3	4
01110		0	4	3
10101			0	3
11011				0

Пример 7.3. Определить корректирующую способность кода, имеющего следующие разрешенные комбинации: 00000; 01110; 10101; 11011.

Решение. Корректирующая способность кода определяется кодовым расстоянием. Составим матрицу расстояний между кодовыми комбинациями (табл. 19).

Как видно из матрицы, кодовое расстояние $d_{\text{мин}} = 3$. Следовательно, данный код способен:

1. Обнаруживать двукратные ошибки.
2. Устранять однократные ошибки.
3. Устранять и обнаруживать однократные ошибки.

Пример 7.4. Определить значность кода n, обеспечивающего исправление всех однократных ошибок при количестве разрешенных комбинаций $N = 8$.

Решение. Вычисление может быть произведено по формуле

$$N \leqslant \frac{2^n}{1+n}.$$

По данным табл. 17 при $N = 8$ требуется значность кода $n=6$.

Пример 7.5. Передается 100 команд простым кодом. Известно, что вероятность искажения одного символа кодовой комбинации под воздействием помех $p_э = 10^{-3}$.

Требуется:

1) определить необходимую значность кода;

2) оценить его помехоустойчивость;

3) оценить помехоустойчивость при трехкратном повторении передачи с групповой и посимвольной проверкой.

Решение. 1) Значность простого кода определяется из соотношения

$$N = 2^n,$$

откуда

$$n = \frac{\lg 100}{\lg 2} = 6{,}6.$$

Округляя до ближайшего целого, получим

$$n = 7.$$

2) Оценка помехоустойчивости кода может быть произведена по формуле (7.38).

Если вероятность искажения одного символа кодовой комбинации равна $p_э$, то при независимости искажений вероятность того, что все n символов кодовой комбинации не будут искажены, равна $(1 - p_э)^n$. Тогда вероятность неправильного приема кодовой комбинации (вероятность искажения комбинации) выразится следующим образом:

$$p_{н.о} = 1 - (1 - p_э)^n \cong np_э = 7 \cdot 10^{-3}.$$

Помехоустойчивость кода

$$S = \lg \frac{1}{p_{н.о}} = \lg \frac{1}{7 \cdot 10^{-3}} = \lg 143 = 2{,}16.$$

3) При трехкратной передаче с групповой проверкой, согласно критерию большинства, условия правильного приема будут следующими:

а) все три кодовые посылки приняты правильно; вероятность такого события равна $(1 - p_э)^{3n}$;

б) из трех посылок две приняты правильно; вероятность такого события равна $C_3^{\prime}(1 - p_э)^{2n}[1 - (1 - p_э)^n]$, где $C_3^{\prime}$ — число случаев,

когда из трех кодовых комбинаций одна принята с ошибкой; $(1-p_э)^{2n}$ — вероятность правильного приема двух посылок, $[1-(1-p_э)^n]$ — вероятность неправильного приема одной посылки.

Таким образом, вероятность ошибочного приема при трехкратном повторении с групповым контролем

$$p_{н.о} = 1 - \{(1-p_э)^{3n} + C_3'(1-p_э)^{2n}[1-(1-p_э)^n]\} =$$
$$= 1-(1-p_э)^{2n}\{(1-p_э)^n + C_3'[1-(1-p_э)^n]\} \cong 1 -$$
$$-[(1-2np_э)(1+2np_э)] = 1-(1-4n^2p_э^2) = 4_n^2p_э^2 = 4\cdot 7^2\cdot 10^{-6} = 2\cdot 10^{-4}.$$

Помехоустойчивость передачи

$$S = \lg\frac{1}{p_{н.о}} = \lg\frac{1}{2\cdot 10^{-4}} = 3{,}7;$$

5) при трехкратной передаче и посимвольном контроле возможны следующие варианты правильного приема одного определенного символа кодовой комбинаций:

а) символ принят правильно во всех трех посылках; вероятность такого события равна $(1-p_э)^3$;

б) символ принят правильно в двух из трех посылок; вероятность такого события равна $C_3'(1-p_э)^2p_э$.

Следовательно, вероятность правильного приема символа кодовой комбинации при трех случаях будет равна $(1-p_э)^3 + C_3'(1-p_э)^2p_э$. Вероятность правильного приема n-значного кода равна $[(1-p_э)^3 + C_3'(1-p_э)^2p_э]^n$. Таким образом, вероятность ошибочного приема

$$p_{н.о} = 1 - [(1-p_э)^3 + C_3'(1-p_э)^2p_э]^n \cong 1 -$$
$$-(1-6p_э^2)^n \cong 6np_э^2 = 42\cdot 10^{-6}.$$

Помехоустойчивость передачи

$$S = \lg\frac{1}{p_{н.о}} = \lg\frac{1}{42\cdot 10^{-6}} = 4{,}38.$$

Следовательно, расчеты показали, что повторение передачи обеспечивает повышение помехоустойчивости передачи, причем посимвольная проверка более эффективна, чем групповая.

Пример 7.6. Необходимо обеспечить передачу 8-ми команд кодом с четным числом единиц. Определить необходимую значность кода, рассчитать избыточность и помехоустойчивость кода при условии, что вероятность искажения одного символа комбинации равна $p_э = 10^{-3}$.

Решение. 1) Код с четным числом единиц строится путем добавления к k информационным символам комбинации одного избыточного (контрольного) символа. Причем в тех случаях, когда информационная часть комбинации содержит четное число единиц, дополнительный символ представляется нулем, при нечетном числе единиц в информационной группе дополнительный символ представляется единицей.

Значность кода, таким образом,

$$n = k + 1,$$

где

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 8}{\lg 2} = 3.$$

Следовательно,

$$n = 4.$$

Возможные варианты кодовых комбинаций представлены в табл. 20.

Как видно из таблицы, код характеризуется расстоянием $d_{мин} = 2$. Однако благодаря специальной логике построения и проверки кода, он позволяет обнаруживать не только однократные ошибки, но и все ошибки нечетной кратности (в нашем случае однократные и трехкратные);

Таблица 20

№ п. п.	Информационные символы	Проверочный символ
1	000	0
2	001	1
3	010	1
4	011	0
5	100	1
6	101	0
7	110	0
8	111	1

2) избыточность кода

$$r = \frac{\rho}{n} = \frac{1}{4} = 0{,}25;$$

3) помехоустойчивость кода будем оценивать по формулам (7.38) и (7.40).

Вероятность неправильного приема кодовой комбинации определяется вероятностью появления необнаруживаемых (четных) ошибок.

Пусть в комбинации искажаются два определенных символа, а остальные не искажаются, тогда вероятность такого события равна $p_э^2(1 - p_э)^{n-2}$. Так как таких вариантов будет C_n^2, то вероятность двукратных ошибок

$$p_{ош}^{(2)} = C_n^2 p_э^2 (1 - p_э)^{n-2}.$$

По аналогии вероятность четырехкратных ошибок

$$p_{ош}^{(4)} = C_n^4 p_э^4 (1 - p_э)^{n-4}.$$

Следовательно, суммарная вероятность появления необнаруживаемых ошибок

$$p_{н.о} = p_{ош}^{(2)} + p_{ош}^{(4)} = C_n^2 p_э^2 (1 - p_э)^{n-2} + C_n^4 p_э^4 (1 - p_э)^{n-4}.$$

Так как с повышением кратности вероятность ошибок резко падает, то

$$p_{н.о} \cong C_n^2 p_э^2 (1 - p_э)^{n-2}.$$

Следовательно, помехоустойчивость кода

$$S = \lg \frac{1}{p_{\text{н. о}}} \cong \lg \frac{1}{C_n^2 p_э^2 (1-p_э)^{n-2}} =$$

$$= \lg \frac{2}{n(n-1)\,p_э^2 (1-p_э)^{n-2}} = 6{,}78.$$

Для нахождения коэффициента обнаружения $K_{обн}$ необходимо определять вероятность появления обнаруживаемых ошибок, которая для кода с проверкой на четность

$$p_{\text{о. о}} = C'_n p_э (1-p_э)^{n-1} + C_n^3 p_э^3 (1-p_э)^{n-3} \cong C'_n p_э (1-p_э)^{n-1}.$$

Тогда коэффициент обнаружения

$$K_{обн} = \frac{p_{\text{о. о}}}{p_к} \cong \frac{C'_n p_э (1-p_э)^{n-1}}{1-(1-p_э)^n} \cong \frac{n p_э (1-p_э)^{n-1}}{n p_э} =$$

$$= (1-p_э)^{n-1} = (1-10^{-3})^3 = 0{,}991.$$

Пример 7.7. Необходимо обеспечить передачу 100 команд инверсным кодом. Определить необходимую значность кода, рассчитать избыточность и помехоустойчивость кода при условии, что вероятность искажения одного символа комбинации $p_э = 10^{-3}$.

Таблица 21

№ п.п.	Информационные символы	Проверочные символы
1	0000000	0000000
2	0000001	1111110
3	0000010	1111101
4	0000011	0000011
5	0000100	1111011
6	.	
10	0001001	0001001
15	0001110	1110001

Решение. 1) В основу построения инверсного кода, как отмечалось в 7.13, положен метод повторения исходной комбинации. Причем в тех случаях, когда исходная комбинация содержит четное число единиц, вторая комбинация в точности воспроизводит исходную, если же исходная комбинация содержит нечетное число единиц, то повторение происходит в инвертированном виде.

Количество информационных символов кода должно быть

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 100}{\lg 2} = 6{,}6. \text{ Выбираем } k = 7.$$

Общее количество символов должно в два раза превышать количество информационных символов

$$n = 2,\ k = 14.$$

Отдельные варианты кодовых комбинаций представлены в табл. 21.

2) Избыточность кода

$$r = \frac{\rho}{n} = \frac{7}{14} = 0{,}5.$$

3) Помехоустойчивость кода оценивается по формулам (7.38) и (7.40).

Как видно из табл. 21, код характеризуется расстоянием $d_{мин} = 4$. Однако, благодаря специальной логике построения и проверки кода, он позволяет практически обнаруживать почти все возможные ошибки.

Наиболее вероятным видом не обнаруживаемых ошибок является одновременное искажение двух символов в исходной комбинации и соответствующих им двух символов в повторяемой комбинации.

Вероятность одновременного искажения какой-либо пары символов в исходной комбинации равна $C_{n/2}^2 p_э^2 (1 - p_э)^{n/2-2}$. Вероятность одновременного искажения двух пар соответствующих символов

$$p_{н.о} = [C_{n/2}^2 p_э^2 (1 - p_э)^{n/2-2}]^2 \cong (C_{n/2}^2)^2 p_э^4.$$

Помехоустойчивость кода

$$S = \lg \frac{1}{p_{н.о}} = \lg \frac{1}{(C_{n/2}^2)\ p_э^4} = \lg \frac{1}{(C_7^2)^2 \cdot 10^{-12}} = 9{,}36.$$

Коэффициент обнаружения

$$K_{обн} = \frac{p_{о.о}}{p_к} = 1 - \frac{p_{н.о}}{p_к} \cong 1 - \frac{(C_{n/2}^2)^2 p_э^4}{1 - (1 - p_э)^n} \cong 1 - \frac{(C_{n/2}^2)^2\, p_э^3}{n} \cong$$

$$\cong 1 - \frac{(n-2)\,(C_{n/2}^2)^2\, p_э^3}{8} = 1 - \frac{(14-2)\, 21^2 \cdot 10^{-9}}{8} = 0{,}99999.$$

Пример 7.8. Построить код Хэмминга, обеспечивающий устранение однократных ошибок и предназначенный для передачи 100 команд. Оценить избыточность и помехоустойчивость кода, если вероятность искажения одного символа комбинации $p_э = 10^{-3}$.

Решение. Количество информационных символов кода

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 100}{\lg 2} = 6{,}6.$$

Выбираем $k = 7$. Необходимую значность кода определим из соотношения

$$N \leqslant \frac{2^n}{1+n}.$$

При $N = 100$ значность кода $n = 11$.

Количество проверочных групп определяется количеством избыточных символов

$$\rho = n - k = 4.$$

Так как значность кода $n = 11$, то проверочные группы должны включать в себя следующие символы:

$$S_1 = a_1 \oplus a_3 \oplus a_5 \oplus a_7 \oplus a_9 \oplus a_{11};$$
$$S_2 = a_2 \oplus a_3 \oplus a_6 \oplus a_7 + a_{10} \oplus a_{11};$$
$$S_3' = a_4 \oplus a_5 \oplus a_6 \oplus a_7;$$
$$S_4 = a_8 \oplus a_9 \oplus a_{10} \oplus a_{11}.$$

Контрольные символы: a_1, a_2, a_4, a_8.

При составлении кодовых комбинаций вначале проставляются информационные символы, а затем контрольные так, чтобы сумма единиц в каждой контрольной группе была четной.

Отдельные варианты кодовых комбинаций представлены в табл. 22:

Таблица 22

№ п.п.	Символы кода										
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
1	1	1	0	0	0	0	0	1	0	0	1
2	1	0	0	0	0	0	0	0	0	1	1
3	1	0	0	0	0	0	0	1	1	0	0
4	0	1	0	0	0	0	0	0	1	0	1
5	0	0	0	0	0	0	0	1	1	1	1

2) избыточность кода

$$r = \frac{\rho}{n} = \frac{4}{11} = 0{,}36;$$

3) безошибочный прием комбинации в коде Хэмминга будет иметь место в двух случаях: в случае отсутствия каких-либо ошибок и в случае наличия однократных ошибок. Соответствующие этим случаям вероятности будут равны $(1 - p_э)^n$ и $C'_n p_э (1 - p_э)^{n-1}$. Тогда вероятность ошибочного приема кодовой комбинации будет

$$p_к = 1 - [(1 - p_э)^n + C'_n p_э (1 - p_э)^{n-1}].$$

Так как код может быть использован для исправления однократных или обнаружения двукратных ошибок, то вероятность появления обнаруживаемых ошибок

$$p_{н.\,о} \cong C_n^3 p_э^3 (1 - p_э)^{n-3}.$$

Следовательно, помехоустойчивость кода

$$S = \lg \frac{1}{p_{н.\,о}} = \lg \frac{1}{C_n^3 p_э^3 (1 - p_э)^{n-3}} \cong \lg \frac{1}{C_n^3 p_э^3} = \lg \frac{1}{C_n^3 10^{-9}} = 6{,}78.$$

Коэффициент обнаружения

$$K_{обн} = 1 - \frac{p_{н.\,о}}{p_к} = 1 - \frac{C_n^3 p_э^3 (1 - p_э)^{n-3}}{1 - [(1 - p_э)^n + C'_n p_э (1 - p_э)^{n-1}]} \cong$$

$$\cong 1 - \frac{C_n^3 p_э^3 (1 - p_э)^{n-3}}{n p_э [1 - (1 - (1 - p_э)^{n-1}]} \cong 1 - \frac{(n-2) p_э (1 - p_э)^{n-3}}{6} =$$

$$= 1 - \frac{(11 - 2) \cdot 10^{-3} (1 - 10^{-3})^8}{6} = 0{,}9985.$$

Пример 7.9. Закодировать простую информационную группу $G(x) = 1011$ циклическим кодом, обеспечивающим обнаружение двукратных или устранение однократных ошибок.

Решение. По заданному количеству информационных символов $k = 4$ определяем значность кода, пользуясь соотношением

$$N = 2^k \leqslant \frac{2^n}{1+n}.$$

Получим $n = 7$.

Для построения циклического кода необходимо выбрать образующий полином $P(x)$ степени $\rho = n - k = 3$, который должен входить в качестве сомножителя для различия двучлена $x^n + 1 = x^{2^z-1}$. В нашем случае этот двучлен имеет вид $x^7 + 1$. Составляющие его сомножители должны быть неприводимыми полиномами, степени которых являются делителями числа $z = 3$. Следовательно, сомножителями двучлена $x^7 + 1$ должны быть неприводимые полиномы первой и третьей степеней.

Пользуясь таблицами неприводимых полиномов в [42, 43], получим

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Выбираем в качестве образующего полинома сомножитель

$$P(x) = x^3 + x^2 + 1.$$

Кодирование осуществляем первым способом.

Для этого исходную кодовую комбинацию $G(x)$ умножаем на $x^{n-k} = x^3$

$$x^{n-k}G(x) = x^3(x^3 + x + 1) = x^6 + x^4 + x^3.$$

Определяем остаток $R(x)$ от деления $x^{n-k}G(x)$ на образующий многочлен $P(x)$

$$\begin{array}{r|l}
x^6 + x^4 + x^3 & x^3 + x^2 + 1 \\
\cline{2-2}
x^6 + x^5 + x^3 & x^3 + x^2 \\
\cline{1-1}
x^5 + x^4 & \\
x^5 + x^4 + x^2 & \\
\cline{1-1}
x^2 &
\end{array}$$

Остаток $R(x) = x^2$.

Следовательно, полином $F(x)$ циклического кода в соответствии с (7.46) будет иметь вид

$$F(x) = x^{n-k}C(x) + R(x) = x^6 + x^4 + x^3 + x^2 = 1011100.$$

Пример 7.10. Получено сообщение циклическим кодом $F^*(x) = x^6 + x^4 + x^3 + x^2$. Проверить декодированием наличие ошибок в принятой комбинации, если образующий полином $P(x) = x^3 + x^2 + 1$.

Решение. Декодирование осуществляется делением полинома полученной комбинации на образующий полином

$$\begin{array}{r|l}
x^6 + x^4 + x^3 + x^2 & x^3 + x^2 + 1 \\
\cline{2-2}
x^6 + x^5 + x^3 & x^3 + x^2 \\
\cline{1-1}
x^5 + x^4 + x^2 & \\
x^5 + x^4 + x^2 & \\
\cline{1-1}
0 &
\end{array}$$

Остаток от деления $R(x)=0$. Следовательно, комбинация принята без искажений.

Пример 7.11. Получена комбинация $F^*(x)=x^6+x^4+x^2+1$, закодированная циклическим кодом. Образующий полином $P(x)=$ $=x^3+x^2+1$. Проверить наличие ошибок в кодовой комбинации.

Решение. Делим полином полученной комбинации на образующий полином

$$\begin{array}{l|l}
x^6+x^4+x^2+1 & x^3+x^2+1 \\
\underline{x^6+x^5+x^3} & \overline{x^3+x^2+1} \\
x^5+x^4+x^3+x^2+1 & \\
\underline{x^5+x^4+x^2} & \\
x^3+1 & \\
\underline{x^3+x^2+1} & \\
\quad x^2 &
\end{array}$$

Остаток $R(x)=x^2\neq 0$. Следовательно, комбинация принята с ошибками.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие критерии применяются для оценки помехоустойчивости систем передачи информации с непрерывными сигналами?
2. Какие критерии применяются для оценки помехоустойчивости систем передачи информации с дискретными сигналами?
3. Перечислите известные способы повышения помехоустойчивости передачи.
4. Что такое системы с повторением передачи? Какие известны разновидности этих систем?
5. Что понимается под системой с обратной связью?
6. Что понимается под системой с информационной обратной связью?
7. Что понимается под системой с решающей обратной связью?
8. Сравните степень загруженности обратного канала в системах с информационной и решающей обратной связью.
9. Сравните эффективность систем с решающей обратной связью, использующих обнаруживающие ошибки коды, с системами без обратной связи, использующими коды с исправлением ошибок.
10. Сравните по помехоустойчивости амплитудную, частотную и кодоимпульсную модуляции.
11. Какие коды называются корректирующими?
12. В чем сущность помехоустойчивого кодирования?
13. Что понимается под значностью и весом кодовой комбинации?
14. Как определяется расстояние между кодовыми комбинациями?
15. Что такое вектор ошибки?
16. Что такое вес вектора ошибки?
17. Что понимается под кодовым расстоянием?
18. Какова связь корректирующей способности кода с кодовым расстоянием?
19. Что представляет собой геометрическая модель комбинаций двоичного кода?
20. Как интерпретируются на геометрической модели кратность ошибок, вес кодовой комбинации и расстояние между кодовыми комбинациями?
21. Какими показателями можно оценивать качество корректирующих кодов?
22. Что понимается под непрерывными и блочными кодами?
23. Что понимается под разделимыми блочными кодами?

24. Что понимается под систематическими и несистематическими блочными кодами?
25. Какие разновидности систематических блочных кодов известны?
26. Какова методика построения кода Хэмминга?
27. В чем заключается логика декодирования Хэмминга?
28. Каковы основные свойства циклических кодов?
29. Какие известны способы построения циклических кодов?
30. В чем заключается методика декодирования циклических кодов?
31. В чем заключается логика построения инверсного кода?
32. Какими корректирующими свойствами обладает инверсный код?

Глава VIII

ЭЛЕМЕНТЫ ТЕОРИИ ОПТИМАЛЬНОГО ПРИЕМА И СТАТИЧЕСКИХ РЕШЕНИЙ

Задача оптимального приема состоит в использовании избыточности, а также имеющихся сведений о свойствах полезного сигнала, помехи и канала для увеличения вероятности правильного приема [37].

Вследствие того что на вход приемника поступает сумма полезного сигнала и помехи, вероятность правильного приема будет определяться отношением полезного сигнала к помехе. Для повышения вероятности правильного приема должна быть произведена предварительная обработка принятого сигнала, обеспечивающая увеличение отношения сигнал/помеха. Таким образом, приемник должен содержать два основных элемента (рис. 8.1): фильтр *Ф*, обеспечивающий улучшение отношения сигнал/помеха, и решающее устройство РУ, выполняющее главные функции приема (обнаружения, различения или восстановления сигналов).

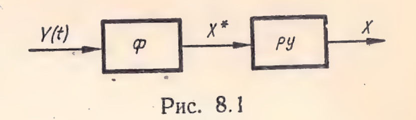


Рис. 8.1

Известны следующие методы фильтрации, обеспечивающие улучшение соотношения сигнал/помеха:

частотная фильтрация;
метод накопления;
корреляционный метод;
согласованная фильтрация.

Все эти методы основаны на использовании различий свойств полезного сигнала и помехи.

8.1. Частотная фильтрация

Идея частотной фильтрации основана на отличии спектров полезного сигнала и помехи. При этом используются линейные частотные фильтры, позволяющие подавлять помеху и улучшать тем самым соотношение сигнал/помеха. Параметры фильтра

определяются спектральными характеристиками сигнала и помехи. На практике наиболее часто встречаются следующие случаи:

а) на вход приемного устройства поступает узкополосный сигнал и широкополосная помеха. Этот случай представлен на рис. 8.2, где $S_x(\omega)$ — спектральная плотность сигнала и $S_\xi(\omega)$ — спектральная плотность помехи.

В этом случае в тракт приемного устройства включается узкополосный фильтр с полосой пропускания $\Delta\omega_x$;

б) на вход приемника поступает широкополосный сигнал и узкополосная помеха (рис. 8.3). В таких случаях в тракт приемника включается фильтр, обеспечивающий подавление помехи в полосе $\Delta\omega_\xi$;

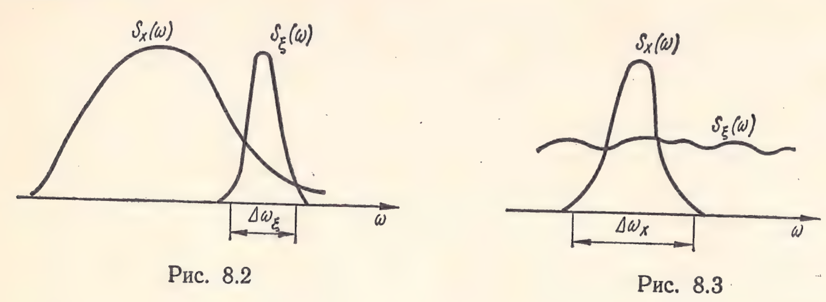


Рис. 8.2

Рис. 8.3

в) на вход приемника поступает периодический сигнал и широкополосная помеха.

Как известно, периодический сигнал имеет дискретный частотный спектр. В этом случае в приемное устройство нужно включать набор фильтров (гребенчатые фильтры), пропускающие лишь дискретные частоты периодического сигнала.

Рассмотрим более детально случай, когда полезный сигнал является гармоническим, а помеха типа белого шума. Для выделения полезного сигнала в этом случае должен быть использован узкополосный фильтр, настроенный на частоту сигнала. Отношение мощности сигнала к мощности помехи на выходе фильтра при этом

$$\left(\frac{P_x^*}{P_\xi}\right)_{\text{вых}} = \frac{P_{x\ \text{вх}}}{P_0 \Delta\omega_\text{ф}} = \frac{P_{x\ \text{вх}}}{2\pi \Delta f_\text{ф} P_0}, \tag{8.1}$$

где P_0 — средняя мощность помехи, приходящаяся на единицу полосы; $\Delta\omega_\text{ф}$ — полоса пропускания фильтра.

Как видно из выражения (8.1), отношение $\frac{P_x^*}{P_\xi}$ можно сделать как угодно большим за счет уменьшения полосы пропускания фильтра $\Delta f_\text{ф}$.

В реальных условиях полезный сигнал поступает лишь в течение отдельного времени T_x. Поэтому полезный сигнал в действительности представляет собой отрезок синусоиды, спектр которого

неограничен. Известно, что практическая ширина спектра такого сигнала связана с его длительностью соотношением

$$\Delta f_x T_x = \mu, \tag{8.2}$$

где μ — постоянная, зависящая от формы сигнала. Обычно принимается $\mu \cong 1$.

Длительность сигнала T_x должна быть выбрана такой, чтобы его спектр был не шире полосы пропускания фильтра

$$\Delta f_x \leqslant \Delta f_ф.$$

Подставляя в (8.1) из (8.2) вместо $\Delta f_ф$ величину Δf_x, получаем

$$\left(\frac{P_x^*}{P_\xi}\right)_{вых} = \frac{P_{x\,вх}}{2\pi P_0 \mu} T_x.$$

Формула (8.3) показывает, что увеличение отношения сигнал/помеха достигается за счет увеличения длительности сигнала T_x, т. е. времени наблюдения.

Таким образом, при частотной фильтрации улучшение отношения сигнал/помеха окупается ценой увеличения времени передачи.

8.2. Метод накопления

Метод накопления применим в том случае, если полезный сигнал в течение времени приема постоянен или является периодической функцией. Метод состоит в многократном повторении сигнала и суммировании отдельных его реализаций в приемном устройстве. Как будет показано ниже, такой способ обработки сигнала обеспечивает при определенных условиях улучшение отношения сигнал/помеха.

Пусть в течение времени передачи T_x сигнал постоянен и равен a. На сигнал воздействует аддитивная помеха ξ.

Последовательность отсчетов можно представить в виде

$$\begin{gathered} y_1 = a + \xi_1; \\ y_2 = a + \xi_2; \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ y_n = a + \xi_n, \end{gathered}$$

где ξ_i — значение помехи в момент i-го отсчета.

На выходе накопителя будет сигнал

$$Y = \sum_{i=1}^{n} (a + \xi_i) = n \cdot a + \sum_{i=1}^{n} \xi_i,$$

где n — количество отсчетов за время передачи.

Отношение мощностей сигнала и помехи на выходе накопителя

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вых}} = \frac{(na)^2}{D\left[\sum\limits_{i=1}^{n} \xi_i\right]},$$

где $D\left[\sum\limits_{i=1}^{n} \xi_i\right]$ — дисперсия помехи на выходе накопителя.

В предположении, что значения ξ_i некоррелированы, дисперсия суммы отсчетов ξ_i равна сумме дисперсий отсчетов

$$D\left[\sum_{i=1}^{n} \xi_i\right] = \sum_{i=1}^{n} D(\xi_i).$$

Считая помеху стационарным случайным процессом, получим

$$D\left[\sum_{i=1}^{n} \xi_i\right] = nD(\xi).$$

Следовательно, отношение мощностей сигнала и помехи на выходе накопителя можно представить в виде

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вых}} = \frac{(na)^2}{nD(\xi)} = n\frac{a^2}{D(\xi)} = n\left(\frac{P_x}{P_\xi}\right)_{\text{вх}}. \qquad (8.4)$$

Таким образом, при перечисленных выше условиях, в результате n-кратного отсчета, отношение мощностей сигнала и помехи увеличивается в n раз. Временной интервал между отдельными отсчетами должен быть больше интервала корреляции помехи τ_0[1]. В противном случае выигрыш за счет накопления будет меньше значения, даваемого выражением (8.4).

За счет увеличения числа отсчетов n, т. е. времени передачи T_x, можно сколь угодно увеличивать отношение сигнал/помеха.

Если сигнал представляет периодическую функцию времени, то отсчеты нужно производить через интервалы, равные или кратные периоду этой функции. В таких случаях метод носит название метода синхронного накопления. Эффект накопления такой же, как и в случае постоянного сигнала.

Эффект накопления можно осуществить также за счет интегрирования входного сигнала в течение определенного времени T_x. Такой метод получил наименование интегрального приема.

Интегральный прием целесообразно применять в случае, когда полезный сигнал постоянен (или квазипостоянен).

---

[1] $\tau_0 = \int\limits_{-\infty}^{\infty} |\rho(\tau)|\, d\tau$. Геометрически интервал корреляции определяется шириной основания прямоугольника единичной высоты, площадь которого равна площади, ограниченной кривой коэффициента корреляции и осью абсцисс.

В случае поступления на вход интегратора суммы постоянного сигнала a и помехи ξ выходной сигнал

$$Y_{\text{и}}(t) = \int_0^{Tx} [a + \xi(t)]\, dt + a \cdot T_x + \int_0^{T_x} \xi(t)\, dt.$$

Отношение мощностей сигнала и помехи на выходе интегратора

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вых}} = \frac{(aT_x)^2}{D\left[\int\limits_0^{T_x} \xi(t)\ dt\right]}.$$

В соответствии с теоремой Котельникова, непрерывный случайный сигнал $\xi(t)$ можно заменить совокупностью отсчетов в моменты времени, отстоящие друг от друга на интервалы, равные интервалу корреляции помехи τ_0. Следовательно, можно записать, что

$$\int_0^{T_x} \xi(t)\, dt \cong \sum_{i=1}^{n} \xi(t_i)\, \tau_0,$$

где

$$n = \frac{T_x}{\tau_0}.$$

Тогда дисперсия помехи на выходе накопителя

$$D\left[\int_0^{T_x} \xi(t)\, dt\right] \cong D\left[\sum_{i=1}^{n} \xi(t_i)\, \tau_0\right] = \tau_0^2 D\left[\sum_{i=1}^{n} \xi(t_i)\right] = \tau_0^2 nD(\xi) = T_x \tau_0 D(\xi).$$

Следовательно, отношение мощностей сигнала и помехи на выходе интегратора определится соотношением

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вых}} \cong \frac{(a \cdot T_x)^2}{T_x \tau_0 D(\xi)} = \frac{T_x}{\tau_0} \frac{a^2}{D(\xi)} = \frac{T_x}{\tau_0} \left(\frac{P_x}{P_\xi}\right)_{\text{вх}}. \qquad (8.5)$$

Таким образом, интегральный прием обеспечивает увеличение отношения мощностей сигнала и помехи в $\left(\frac{T_x}{\tau_0}\right)$ раз. Помехоустойчивость интегрального приема будет тем выше, чем больше время интегрирования T_x и чем более высокочастотна помеха (т. е. чем меньше интервал корреляции помехи τ_0).

8.3. Корреляционный метод

Сущность метода заключается в использовании различия между корреляционными функциями сигнала и помехи. Данный метод эффективен лишь в случае приема периодических или квазипериодических сигналов.

Рассмотрим сущность метода на примере, когда полезный сигнал является периодическим, а помеха — типа белого гауссова шума.

В приемном устройстве определяется корреляционная функция поступающей на вход суммы полезного сигнала и помехи

$$K_y(\tau) = \frac{1}{T_1}\int\limits_{-\frac{T_1}{2}}^{\frac{T_1}{2}} [X(t) + \xi(t)][X(t+\tau) + \xi(t+\tau)]\,dt =$$

$$= \frac{1}{T_1}\int\limits_{-\frac{T_1}{2}}^{\frac{T_1}{2}} X(t)X(t+\tau)\,dt + \frac{1}{T_1}\int\limits_{-\frac{T_1}{2}}^{\frac{T_1}{2}} X(t)\xi(t+\tau)\,dt +$$

$$+ \frac{1}{T_1}\int\limits_{-\frac{T_1}{2}}^{\frac{T_1}{2}} \xi(t)X(t+\tau)\,dt + \frac{1}{T_1}\int\limits_{-\frac{T_1}{2}}^{\frac{T_1}{2}} \xi(t)\xi(t+\tau)\,dt =$$

$$= K_{xx}(\tau) + K_{x\xi}(\tau) + K_{\xi x}(\tau) + K_{\xi\xi}(\tau). \qquad (8.6)$$

В полученном выражении $K_{x\xi}(\tau)$ и $K_{\xi x}(\tau)$ есть взаимные корреляционные функции сигнала и помехи, а $K_{xx}(\tau)$ и $K_{\xi\xi}(\tau)$ — автокорреляционные функции сигнала и помехи соответственно.

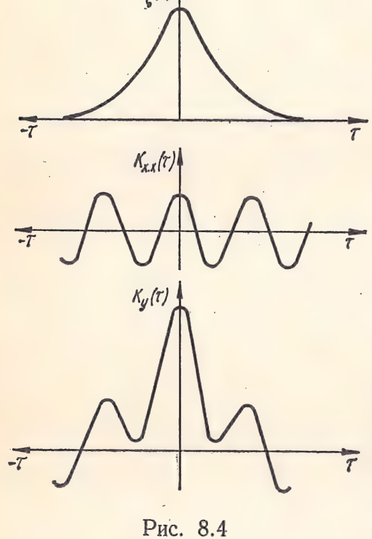


Рис. 8.4

Поскольку передаваемый сигнал и помеха статистически независимы, то

$$K_{x\xi}(\tau) = K_{\xi x}(\tau) = 0.$$

Следовательно, выражение (8.6) примет вид

$$K_y(\tau) = K_{xx}(\tau) + K_{\xi\xi}(\tau). \qquad (8.7)$$

Из выражения (8.7) видно, что корреляционная функция смеси сигнала и помехи равна сумме автокорреляционных функций сигнала и помехи. Как известно, корреляционная функция периодического сигнала является периодической функцией аргумента (см. параграф 3.11, пример 3.14). Функция $K_{\xi\xi}(\tau)$ с увеличением τ стремится к нулю и при $\tau \geqslant \tau_0$ практически равна нулю (рис. 8.4). Следовательно, выбирая такое время τ, при котором значением $K_{\xi\xi}(\tau)$ можно пренебречь, мы обеспечим тем самым получение функции $K_y(\tau)$, отображающей полезный сигнал, т. е. выделение полезного сигнала из смеси полезного сигнала с помехой.

Для уточнения того, от каких факторов зависит время, затрачиваемое для выделения полезного сигнала при корреляционном

приеме, выразим корреляционные функции $K_{xx}(\tau)$ и $K_{\xi\xi}(\tau)$ через дисперсии и нормированные корреляционные функции $r_{xx}(\tau)$ и $r_{\xi\xi}(\tau)$

$$K_{xx}(\tau) = D(x) \cdot r_{xx}(\tau); \qquad (8.8)$$
$$K_{\xi\xi}(\tau) = D(\xi)\, r_{\xi\xi}(\tau).$$

Подставляя (8.8) и (8.7), найдем

$$K_y(\tau) = D(x)\, r_{xx}(\tau) + D(\xi)\, r_{\xi\xi}(\tau) = D(x)\left[r_{xx}(\tau) + \frac{D(\xi)}{D(x)} r_{\xi\xi}(\tau)\right] =$$
$$= D(x)\left[r_{xx}(\tau) + \frac{P_\xi}{P_x} r_{\xi\xi}(\tau)\right]. \qquad (8.9)$$

Из (8.9) видно, что выбор времени приема зависит от интервала корреляции помехи τ_0 и отношения сигнал/помеха. Действительно, при $\frac{P_\xi}{P_x} \ll 1$ вторым членом можно пренебречь и для выделения полезного сигнала теоретически времени не потребуется. При $\frac{P_\xi}{P_x} \gg 1$ необходимо увеличивать τ, чтобы получить $r_{\xi\xi}(\tau) \cong 0$. Следовательно, в последнем случае для выделения полезного сигнала необходимо дополнительное время, которое должно расти с увеличением отношения $\frac{P_\xi}{P_x}$ и интервала корреляции помехи τ_0.

Общее время, затрачиваемое на корреляционный прием, определяется не только τ, но и временем интегрирования T_1, выбираемым достаточно большим. Практически ограничиваются значением

$$T_1 \geqslant 10\tau_0.$$

Схема корреляционного приемника приведена на рис. 8.5. Схема содержит блок задержки БЗ, устройство умножения УУ и интегратор И. Блок задержки осуществляет задержку входного сигнала на время τ. Устройство умножения осуществляет перемножение поступающей на вход приемника функции $Y(t)$ и функции $Y(t+\tau)$, снимаемой с блока задержки. Интегратор усредняет полученное произведение за время T_1.

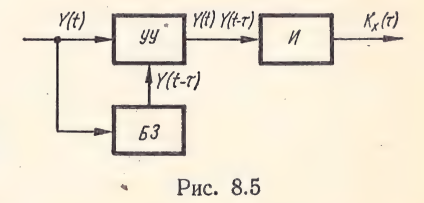


Рис. 8.5

8.4. Согласованная фильтрация

Согласованные фильтры предназначены для выделения сигналов известной формы на фоне шумов. Критерием оптимальности таких фильтров является получение на выходе максимально возможного отношения амплитудного значения сигнала к действующему значению помехи.

Показано [27, 46], что функция передачи таких фильтров должна определяться выражением

$$K(j\omega) = C\frac{S_x^*(j\omega)}{G_\xi(\omega)}\exp\{-j\omega T\}, \tag{8.10}$$

где $S_x^*(j\omega)$ — комплексно-сопряженная функция спектра сигнала; $G_\xi(\omega)$ — энергетический спектр помехи на входе фильтра; T — момент времени наблюдения за сигналом на выходе фильтра, в который достигается максимум отношения сигнал/помеха; C — коэффициент пропорциональности.

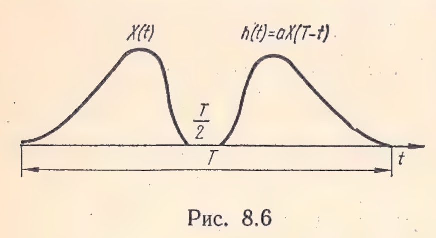


Рис. 8.6

Для случая, когда помеха является белым шумом, формула (8.10) принимает вид

$$K(j\omega) = BS_x^*(j\omega)\exp\{-j\omega T\}, \tag{8.11}$$

где $B = \frac{G}{G_{\xi_0}}$; G_{ξ_0} — энергетический спектр белого шума.

Если полезный сигнал описывается функцией $x(t)$, то переходная функция оптимального фильтра имеет вид

$$h(t) = Bx(T-t), \tag{8.12}$$

т. е. представляет собой зеркальное отображение сигнала относительно временной точки $\frac{T}{2}$ (рис. 8.6).

Реакция на выходе фильтра определяется интегралом свертки

$$Y_\Phi(t) = \int_{-\infty}^{\infty}[x(\tau)+\xi(\tau)]\,h(t-\tau)\,d\tau =$$

$$= B\int_{-\infty}^{\infty}x(\tau)\,x(T-t+\tau)\,d\tau + B\int_{-\infty}^{\infty}\xi(\tau)\,x(T-t+\tau)\,d\tau. \tag{8.13}$$

Формула (8.13) показывает, что выходная реакция фильтра с точностью до постоянного множителя выражает сумму автокорреляционной функции сигнала и взаимно корреляционной функции сигнала и помехи. Следовательно, реакция согласованного фильтра эквивалентна действию корреляционного приемника.

8.5. Сущность основной задачи приема сигналов при наличии помех

Результатом воздействия помех является частичная или полная потеря информации, переносимой полезным сигналом. Приемное устройство, осуществляя обработку входного сигнала, являющегося суммой полезного сигнала и помехи, должно обеспечить извлече-

ние из принятого сигнала возможно большего количества необходимой информации.

Основная задача приемника состоит в том, чтобы на основании принятой реализации решить наилучшим в каком-то определенном смысле способом, имеется ли данный сигнал в данной реализации (задача обнаружения или различения), или каковы параметры полезного сигнала (задача восстановления). В связи с этим должны быть выработаны критерии, позволяющие по принятому сигналу оптимальным способом решить поставленную задачу.

Задача выбора оптимального способа обработки сигналов и выработки при этом соответствующих критериев составляет содержание теории статистических решений. Некоторые положения этой теории приводятся ниже, а более детально изложены в [10, 14, 47].

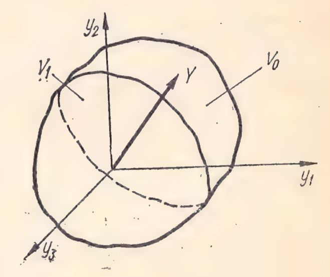


Рис. 8.7

С целью наглядного представления положений теории статистических решений введены геометрические понятия пространства принимаемого сигнала (пространства наблюдений).

Пусть отсчеты принимаемого сигнала, являющегося суммой полезного сигнала и помехи, осуществляются в дискретные моменты времени $t_1, t_2, \ldots, t_n$. Отсчетные значения принятого сигнала $y_1, y_2, \ldots, y_n$ называют выборочными значениями, а их совокупность — выборкой. Число n выборочных значений называют размером (или объемом) выборки.

Совокупность выборочных значений представляют геометрически в виде радиус-вектора Y в n-мерном пространстве, где $y_1, y_2 \ldots, y_n$ — координаты конца вектора. Так как величины $y_1, y_2, \ldots, y_n$ случайны, то вектор Y также является случайным вектором. Множество возможных значений вектора Y составляет пространство наблюдений V. На рис. 8.7 показан случай трехмерного пространства ($n = 3$). Общая вероятность попадания конца вектора Y в произвольную точку пространства V

$$\int\limits_V w(y)\,dy = 1. \tag{8.14}$$

Но аналогии вводят понятия вектора полезного сигнала и вектора помех и соответственно им понятие пространства полезного сигнала и пространства помех.

После нахождения вектора принятого сигнала Y мы не можем однозначно судить о векторе полезного сигнала X. Речь может идти только об апостериорной плотности вероятности $w(X/Y) = w(x_1, x_2, \ldots, x_n/y_1, y_2, \ldots, y_n)$, т. е. условной плотности вероятности X, если задан вектор Y.

Вычисление апостериорной плотности вероятности можно выполнить с помощью формулы Байеса

$$w(X/Y)=\frac{w(X)\,w(Y/X)}{w(Y)}, \tag{8.15}$$

где $w(X)$ — априорная плотность вероятности вектора X; $w(Y)$ — безусловная плотность вероятности вектора Y; $w(Y/X)$ — условная плотность вероятности Y, если задан X.

Безусловная плотность вероятности $w(Y)$ определяется соотношением

$$w(Y)=\int\limits_{V_x} w(X)\,w(Y/X)\,dX, \tag{8.16}$$

где V_x обозначает, что интегрирование осуществляется в пространстве сигнала X.

Подставляя значение $w(Y)$ из (8.16) в (8.15), получим

$$w(X/Y)=\frac{w(X)\,w(Y/X)}{\int\limits_{V_x} w(X)\,w(Y/X)\,dx}. \tag{8.17}$$

Если вектор X может иметь конечное число возможных значений $x_1, x_2, \ldots, x_r$ с априорными вероятностями $p(x_1), p(x_2), \ldots, p(x_r)$, то формула (8.17) принимает вид

$$p(X/Y)=\frac{p(X)\,w(Y/X)}{w(Y)}=\frac{p(X)\,w(Y/X)}{\sum\limits_{j=1}^{r} p(x_j)\,w(Y/x_j)}, \tag{8.18}$$

где $p(X/Y)$ — апостериорная вероятность вектора X, если задан вектор Y; $p(X)$ — априорная вероятность вектора X.

Следовательно, для нахождения искомой апостериорной вероятности (или плотности вероятности) необходимо знать $P(X)$ или $w(X)$, т. е. априорные характеристики полезного сигнала, и $w(Y/X)$, определяемые априорными характеристиками полезного сигнала и помехи, а также характером их композиции.

Для определения апостериорных вероятностей $p(X/Y)$ или плотностей вероятностей $w(X/Y)$ необходимо знать $w(Y/X)$, которая при заданном значении Y будет зависеть только от X

$$w(Y/X)=L(X). \tag{8.19}$$

Функция $L(X)$ называется функцией правдоподобия. В зависимости от того, является ли X дискретной или непрерывной величиной, функция правдоподобия $L(X)$ может принимать конечное или бесконечное множество значений.

Рассмотрим основные критерии, используемые при решении задачи оптимального приема. Начнем с простейшей задачи — задачи обнаружения сигналов.

8.6. Обнаружение сигнала

Задача обнаружения, как отмечалось, состоит в том, чтобы в результате обработки принятого сигнала Y установить, содержится ли в нем полезный сигнал X или нет.

Пусть принимаемый сигнал является суммой полезного сигнала и помехи

$$y(t) = x(t) + \xi(t).$$

Полезный сигнал может принимать два значения: x_1 и x_0 с априорными соответственно вероятностями $p(x_1)$ и $p(x_0)$. Так как сигнал X наверняка имеет одно из этих двух значений, то справедливо соотношение

$$p(x_1) + p(x_0) = 1. \tag{8.20}$$

Таким образом, возможны две взаимно исключающие (альтернативные) гипотезы: в принятом сигнале содержится полезный сигнал (гипотеза H_1) и отсутствует полезный сигнал (гипотеза H_0). Решающее устройство приемника по данным выборки должно установить, какая из этих гипотез является истинной.

В геометрической интерпретации поставленная задача может быть сформулирована следующим образом. Пространство принятых сигналов V условно разбивается на две части (рис. 8.7): область v_1, соответствующую принятию гипотезы H_1 о том, что $X = x_1$, и область v_0, соответствующую принятию гипотезы H_0 о том, что $X = x_0$. Это значит, что если вектор принятого сигнала окажется в пределах области v_1, то принимается гипотеза H_1. Если же вектор сигнала Y окажется в области v_0, то принимается гипотеза H_0.

В этих условиях могут иметь место два значения апостериорной вероятности $p(X/Y)$: $p(X_1/Y)$ — условная вероятность наличия полезного сигнала X при данном значении выборки Y, $p(X_0/Y)$ — условная вероятность отсутствия x при данном значении выборки Y.

Аналогично можно рассматривать два значения функции правдоподобия $L(X)$: $L(x_1) = w(Y/X)$ — условная плотность вероятности выборки Y при наличии полезного сигнала X; $L(x_0) = w(Y/x_0)$ — условная плотность вероятности выборки Y при отсутствии X.

Отношение функций правдоподобия

$$\lambda = \frac{L(x_1)}{L(x_0)} = \frac{w(Y/x_1)}{w(Y/x_0)} \tag{8.21}$$

принято называть отношением правдоподобия.

Для выбора гипотезы H_1 или H_0 должно быть взято за основу определенное правило принятия решений.

Выбор правила принятия решения в математическом отношении сводится к оптимальному разбиению пространства принимаемых сигналов V на области v_1 и v_0.

Для того чтобы выбрать то или иное правило принятия решения, необходимо руководствоваться определенными критериями.

Критерий максимума правдоподобия. Этот критерий формулируется следующим образом: наиболее правдоподобно то значение параметра X, для которого функция правдоподобия $L(X)$ максимальна.

В соответствии с этим критерием в случае двухальтернативной ситуации (обнаружение сигнала) сравнивается два значения функции правдоподобия — $L(x_1)$ и $L(x_0)$ и принимается та гипотеза, которой соответствует большее значение функции правдоподобия. Если, например, $L(x_1) > L(x_0)$, то принимается гипотеза H_1. Если же $L(x_1) \leqslant L(x_0)$, то принимается гипотеза H_0.

Этот критерий можно записать в следующем виде через отношение правдоподобия:

$$\text{если } \lambda = \frac{L(x_1)}{L(x_0)} > 1, \text{ то } x = x_1; \tag{8.22}$$

$$\text{при } \lambda = \frac{L(x_1)}{L(x_0)} \leqslant 1, \text{ то } x = x_0.$$

Таким образом, в соответствии с данным критерием методика принятия решения сводится к следующему: вычисляются функции правдоподобия $L(x_1)$ и $L(x_0)$, определяется отношение правдоподобия λ, и в зависимости от того, больше, равно или меньше λ единицы принимается соответствующая гипотеза.

Практическое достоинство данного критерия заключается в том, что при его применении не требуется знания априорных вероятностей $p(x_1)$ и $p(x_0)$ сигнала X.

Критерий максимума апостериорной вероятности. По этому критерию при полученном значении выборки Y принимается та гипотеза, при которой апостериорная вероятность $p(X/Y)$ максимальна.

Для случая двухальтернативной ситуации сравниваются два значения апостериорной вероятности $p(x_1/Y)$ и $p(x_0/Y)$. Обычно рассматривается отношение этих величин и правило принятия решения записывается в виде:

$$\text{если } \frac{p(x_1/Y)}{p(x_0/Y)} > 1, \text{ то } X = x_1;$$

$$\text{если } \frac{p(x_1/Y)}{p(x_0/Y)} \leqslant 1, \text{ то } X = x_0. \tag{8.23}$$

Используя формулу Байеса (8.18), выразим отношение апостериорных вероятностей через отношение функций правдоподобия

$$\frac{p(x_1/Y)}{p(x_0/Y)} = \frac{p(x_1)\,L(X_1)}{p(x_0)\,L(X_0)} = \frac{p(x_1)}{p(x_0)}\lambda. \tag{8.24}$$

Тогда критерий максимума апостериорной вероятности (8.23) может быть следующим образом выражен через отношение правдоподобия:

$$\text{если } \frac{p(x_1)}{p(x_0)}\lambda > 1, \text{ то } X = x_1;$$

$$\text{если } \frac{p(x_1)}{p(x_0)}\lambda \leqslant 1, \text{ то } X = x_0. \tag{8.25}$$

Соотношения (8.25) можно представить в виде:

$$\text{если } \lambda > \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_1;$$

$$\text{если } \lambda \leqslant \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_0, \qquad (8.26)$$

где λ_0 — пороговое значение отношения правдоподобия.

Таким образом, процедура принятия решения согласно критерию максимума апостериорной вероятности такая же, как и согласно критерию максимума правдоподобия. Отличие заключается лишь в том, что в первом случае отношение правдоподобия сравнивается с единицей, а во втором случае — с отношением априорных вероятностей $p(x_0)/p(x_1)$. При наличии априорных данных $p(x_1)$ и $p(x_0)$ целесообразно применять критерий максимума апостериорной вероятности, так как при этом имеется возможность пользоваться дополнительной информацией, позволяющей точнее решить задачу обнаружения сигнала.

Следует заметить, что критерий максимума правдоподобия является оптимальным с информационной точки зрения. Действительно, с точки зрения теории информации наиболее вероятным следует считать то значение параметра X, относительно которого в принятом сигнале Y содержится наибольшее количество информации. Взяв разность количеств информации, содержащихся в сигнале Y относительно x_1 и x_0, получим $I(Y,\ x_1) - I(Y,\ x_0) = [-\log_2 p(x_1) + \log_2 p(x_1/Y)] - [\log_2 p(x_0/Y) -$

$$- \log_2 p(x_2)] = \log_2 \frac{p(x_1/Y)\ p(x_0)}{p(x_0/Y)\ p(x_1)} = \log_2 \frac{w(Y/x_1)}{w(Y/x_0)} \log_2 \lambda.$$

Таким образом, информационный критерий принятия решения сводится к определению двоичного логарифма отношения правдоподобия. Если этот логарифм положителен, то принимается гипотеза H_1 о том, что $X = x_1$; если он отрицателен или равен нулю, то принимается гипотеза H_0 о том, что $X = x_0$.

Критерий идеального наблюдателя (критерий Котельникова). Согласно данному критерию принимается та гипотеза, при которой обеспечивается минимум общей ошибки принятия решения.

При решении задачи обнаружения сигнала могут иметь место ошибки двух родов:

1) при отсутствии полезного сигнала вектор принятого сигнала Y оказывается в области v_1 и принимается в соответствии с этим гипотеза H_1; 2) при наличии полезного сигнала вектор Y оказывается в области v_0 и принимается гипотеза H_0. Первая ошибка называется ошибкой первого рода, или «ложной тревогой». Вторая ошибка называется ошибкой второго рода, или «пропуском сигнала». Количественно ошибки первого и второго рода оцениваются условными вероятностями α и β ошибочных решений о наличии полезного

сигнала, когда в действительности он отсутствует, и об отсутствии сигнала, когда в действительности он имеется

$$\alpha = p(Y \in v_1/x_0) = \int_{v_1} w(Y/x_0)\,dY;$$

$$\beta = p(Y \in v_0/x_1) = \int_{v_0} w(Y/x_1)\,dY. \qquad (8.27)$$

Общая безусловная вероятность ошибочного решения определяется выражением

$$p_{\text{ош}} = p(x_0)\,\alpha + p(x_1)\,\beta. \qquad (8.28)$$

Критерий идеального наблюдателя минимизирует общую ошибку, определяемую выражением (8.28).

Следовательно, условие оптимального решения по критерию идеального наблюдателя имеет вид

$$p_{\text{ош}} = p(x_0)\,\alpha + p(x_1)\,\beta = \text{мин}. \qquad (8.29)$$

Подставим в (8.28) из (8.27) значения ошибок первого и второго рода

$$p_{\text{ош}} = p(x_0)\int_{v_1} w(Y/x_0)\,dY + p(x_1)\int_{v_0} w(Y/x_1)\,dY. \qquad (8.30)$$

Ошибку второго рода можно представить в виде

$$\beta = p(Y \in v_0/x_1) = 1 - p(Y \in v_1/x_1) = 1 - \int_{v_1} w(Y/x_1)\,dY. \quad (8.31)$$

Подставив из (8.31) в (8.30) значение β, получим

$$p_{\text{ош}} = p(x_1) - \int_{v_1} [p(x_1)\,w(Y/x_1) - p(x_0)\,w(Y/x_0)]\,dY. \qquad (8.32)$$

Условие (8.29) будет обеспечено, если интеграл в (8.32) будет максимален. А для этого нужно так выбрать область v_1, чтобы подынтегральная функция была положительной, т. е.

$$p(x_1)\,w(Y/x_1) - p(x_0)\,w(Y/x_0) > 0. \qquad (8.33)$$

Условие (8.33) определяет принадлежность вектора Y области v_1, т. е. выбор гипотезы H_1. Перепишем (8.33) в виде:

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda > \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_1;$$

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda \leqslant \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_0.$$

Таким образом, правила решения, соответствующие критериям идеального наблюдателя и максимума апостериорной вероятности, совпадают. Отличие заключается лишь в исходных условиях.

Критерий Неймана—Пирсона. Данный критерий основан на том, что ошибки первого и второго рода не одинаково опасны, причем ошибка первого рода приводит к таким последствиям, что ее вероятность необходимо ограничить некоторой очень малой величиной. Вторую ошибку желательно при этом обеспечить минимальной.

Исходя из этого, критерий Неймана—Пирсона можно сформулировать следующим образом: наилучшим решением является такое, при котором обеспечивается наименьшая вероятность ошибки второго рода при заданной допустимой вероятности ошибки первого рода.

Итак, согласно критерию Неймана—Пирсона должно быть обеспечено

$$\beta = \int_{v_0} w(Y/x_1)\,dY = \text{мин} \tag{8.34}$$

$$\text{при } \alpha = \int_{v_1} w(Y/x_0)\,dY = \varepsilon, \tag{8.35}$$

где ε — наперед заданная величина.

Задача может быть решена методом Лагранжа отыскания условного экстремума.

Для упрощения задачи целесообразно перейти от многомерной переменной Y к одномерной переменной λ, что можно осуществить с помощью равенств

$$\begin{aligned} w(Y/x_1)\,dY &= w(\lambda/x_1)\,d\lambda; \\ w(Y/x_0)\,dY &= w(\lambda/x_0)\,d\lambda. \end{aligned} \tag{8.36}$$

При таком переходе областям v_1 и v_0 в пространстве V соответствуют области $[0 \div \lambda_0]$ и $[\lambda_0 \div \infty]$ значений λ.

Условные вероятности ошибок первого и второго рода будут при этом представлены в виде

$$\alpha = \int_{v_1} w(Y/x_0)\,dY = \int_{\lambda_0}^{\infty} w(\lambda/x_0)\,d\lambda; \tag{8.37}$$

$$\beta = \int_{v_0} w(Y/x_1)\,dY = \int_{0}^{\lambda_0} w(\lambda/x_1)\,d\lambda. \tag{8.38}$$

Тогда для отыскания условного экстремума должна быть составлена вспомогательная функция

$$F = \lambda \int_{\lambda_0}^{\infty} w(\lambda/x_0)\,d\lambda + \int_{0}^{\lambda_0} w(\lambda/x_1)\,d\lambda.$$

Взяв производную $\frac{\partial F}{\partial \lambda_0}$ и приравняв ее нулю, получим

$$w(\lambda_0/x_1) - \lambda w(\lambda_0/x_0) = 0,$$

или

$$\frac{w(\lambda_0/x_1)}{w(\lambda_0/x_0)} = \lambda.$$

Но из (8.36)

$$\frac{w(\lambda/x_1)}{w(\lambda/x_0)} = \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda$$

и, следовательно,

$$\frac{w(\lambda_0/x_1)}{w(\lambda_0/x_0)} = \lambda_0.$$

Таким образом, данный критерий будет справедлив при $\lambda = \lambda_0$, где пороговое значение λ_0 определяется из равенства

$$\alpha = \int\limits_{\lambda_0}^{\infty} w(\lambda/x_0)\, d\lambda = \varepsilon. \tag{8.39}$$

Итак, правило принятия решения согласно критерию Неймана—Пирсона может быть записано в виде:

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda > \lambda_0, \text{ то } X = x_1;$$

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda \leqslant \lambda_0, \text{ то } X = x_0. \tag{8.40}$$

Критерий минимального риска (критерий Байеса). Этот критерий учитывает не только неравноценность ошибок первого и второго рода, но и те последствия, к которым приводят эти ошибки. Для учета этих последствий введены весовые коэффициенты (коэффициенты цены ошибок) r_{10} и r_{01}, приписываемые соответственно ошибкам первого и второго рода.

Усредненная величина

$$r = r_{10}p(x_0)\,\alpha + r_{01}p(x_1)\,\beta \tag{8.41}$$

получила название риска.

В соответствии с критерием минимального риска правило выбора решения формулируется следующим образом: принимается та гипотеза, при которой обеспечивается минимальный риск:

$$r = \text{мин}. \tag{8.42}$$

Представим (8.41) в виде

$$r = r_{10}p(x_0)\int\limits_{v_1} w(Y/x_0)\,dY + r_{01}p(x_1)\int\limits_{v_0} w(Y/x_1)\,dY =$$

$$= r_{01}p(x_1) - \int\limits_{v_1}[r_{01}p(x_1)\,w(Y/x_1) - r_{10}p(x_0)\,w(Y/x_0)]\,dY. \tag{8.43}$$

Минимум выражения (8.43) будет при условии, если подынтегральная функция положительная

$$r_{01}p(x_1)\,w(Y/x_1) - r_{10}p(x_0)\,w(Y/x_0) > 0.$$

Отсюда получаем следующее правило принятия решения:

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda > \frac{r_{10} p(x_0)}{r_{01} p(x_1)} = \lambda_0, \text{ то } X = x_1; \tag{8.44}$$

$$\text{если } \frac{w(Y/x_1)}{w(Y/x_0)} = \lambda \leqslant \frac{r_{10} p(x_0)}{r_{01} p(x_1)} = \lambda_0, \text{ то } X = x_0.$$

Рассматриваемый критерий наиболее целесообразен экономически, так как обеспечивает минимизацию потерь, обусловленных ошибками в принятии решений. Но он требует максимальной априорной информации, ибо помимо функций распределения $w(Y/X)$ и априорных вероятностей $p(X)$, необходимо также знание весовых коэффициентов r_{10} и r_{01}.

Минимаксный критерий. Минимаксный критерий представляет собой специальный случай критерия минимального риска, когда априорные вероятности $p(x_1)$ и $p(x_0)$ не заданы.

Дело в том, что риск r, получающий наименьшее значение при условии (8.44), зависит от априорных вероятностей $p(x_1)$ и $p(x_0)$. При определенном соотношении этих вероятностей, который мы назовем наихудшим, риск r будет максимален.

Идея минимаксного критерия заключается в том, что обеспечивается минимум риска при наихудшем соотношении априорных вероятностей.

Для определения наихудшего соотношения между $p(x_1)$ и $p(x_0)$ необходимо приравнять нулю производную от правой части (8.43)

Таблица 23

№ п. п.	Наименование критерия	Пороговое значение отношения правдоподобия
1	Максимум правдоподобия	$\lambda_0 = 1$
2	Максимум апостериорной вероятности	$\lambda_0 = \frac{p(x_0)}{p(x_1)}$
3	Идеальный наблюдатель	$\lambda_0 = \frac{p(x_0)}{p(x_1)}$
4	Нейман-Пирсон	λ_0 находится из условия $\int_{\lambda_0}^{\infty} w(\lambda/x_0)\, d\lambda = \varepsilon$
5	Минимальный риск	$\lambda_0 = \frac{r_{10} p(x_0)}{r_{01} p(x_1)}$
6	Минимакс	$\lambda_0^* = \frac{r_{10} p^*(x_0)}{r_{01} p^*(x_1)}$, где $p^*(x_0)$ и $p^*(x_1)$ находятся из условия $\frac{\partial r}{\partial p(x_1)} = 0$

по $p(x_1)$ (или по $p(x_0)$). В результате получается трансцендентное уравнение, обеспечивающее максимум риска. Затем определяется пороговое значение отношения правдоподобия

$$\lambda_0^* = \frac{r_{10} p^*(x_0)}{r_{01} p^*(x_1)}, \tag{8.45}$$

где $p^*(x_0)$ и $p^*(x_1)$ — наиболее неблагоприятные значения априорных вероятностей $p(x_0)$ и $p(x_1)$, полученные из условия $\frac{\partial r}{\partial p(x_1^*)} = 0$.

Таким образом, правило принятия решения для всех рассмотренных критериев одинаково и сводится к сравнению отношения правдоподобия λ с пороговым значением λ_0. Отличие заключается лишь в величине λ_0^*. Для наглядности значения λ_0 для отдельных критериев сведены в табл. 23.

Так как величина λ_0 определяет границу между областями v_1 и v_0 пространства V, то каждый критерий определяет способ разбивки пространства принятого сигнала на области v_1 и v_0. Равенство $\frac{w(Y/x_1)}{w(Y/x_0)} = \lambda_0$ определяет уравнение поверхности раздела областей v_1 и v_0.

8.7. Различение сигналов

При различении сигналов имеет место многоальтернативная ситуация, когда полезный сигнал X может иметь много значений и приемное устройство должно определить, какое именно значение из этого множества имеет место в действительности. Различение многих сигналов в принципиальном отношении мало отличается от случая обнаружения сигнала, т. е. случая различения двух сигналов. В соответствии с этим методы многоальтернативных решений являются обобщением соответствующих методов двухальтернативных решений.

Пусть сигнал X может иметь m возможных значений x_1, x_2, ..., x_m с априорными вероятностями $p(x_1)$, $p(x_2)$, ..., $p(x_m)$ соответственно

$$X = \begin{cases} x_1 \to p(x_1); \\ x_2 \to p(x_2); \\ \quad . \; . \; . \; . \\ x_m \to p(x_m). \end{cases}$$

При этом пространство сигнала V разбивается на m областей v_1, v_2, ..., v_m соответствующих принятию гипотез H_1, H_2, ..., H_m о том, что $X = x_1$, $X = x_2$, ..., $X = x_m$ соответственно. Правила принятия решений и разбивка пространства V на области v_1, v_2, ..., v_m могут производиться в соответствии с любым из критериев, рассмотренных для случая двухальтернативной ситуации и обобщенных на случай многоальтернативной ситуации.

Процедура работы решающего устройства приемника при различении сигналов следующая. По данным выборки Y определяются функции правдоподобия $L(x_1) = w(Y/x_1)$, $L(x_2) = w(Y/x_2)$, $\ldots$, $L(x_m) = w(Y/x_m)$ и вычисляются отношения правдоподобия $\lambda_{ji} = \frac{w(Y/x_j)}{w(Y/x_i)}$ для всех возможных сочетаний пар x_j и x_i. Сравниваются полученные значения отношений правдоподобия с пороговым значением и выбирается такое значение сигнала x_j, для которого все $\lambda_{ji} > \lambda_0$ $(i = 1, 2, \ldots, m)$.

Рассмотрим в качестве примера случай, когда используется критерий минимального риска.

В случае многоальтернативной ситуации ошибки принятия решения заключается в том, что наблюдаемая выборка оказывается в области v_k, в то время, как в действительности сигнал X имеет значение x_j. Цена ошибочных решений учитывается путем введения весовых коэффициентов r_{jk}.

Для заданного значения сигнала x_j средняя величина потерь за счет неправильных решений может быть оценена коэффициентом

$$r_j = \sum_{k=1}^{m} r_{jk} p(Y \in v_k/x_j) = \sum_{k=1}^{m} r_{jk} \int_{v_k} w(Y/x_j)\, dY, \tag{8.46}$$

где $p(Y \in v_k/x_j)$ — условная вероятность попадания выборки Y в область v_k, если в действительности сигнал X равен x_j.

Величины r_j носят название условного риска.

Усредняя условный риск по всем возможным значениям X, получим средний риск

$$r = \sum_{j=1}^{m} p(x_j) r_j = \sum_{j=1}^{m} \sum_{k=1}^{m} p(x_j) r_{jk} p(Y \in v_k/x_j) =$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{m} p(x_i) r_{jk} \int_{v_k} w(Y/x_j)\, dY. \tag{8.47}$$

Критерий минимального риска для случая многоальтернативной ситуации сводится к минимизации функции (8.47)

$$r = \text{мин}. \tag{8.48}$$

Условие (8.48) определяет правила принятия решения, а также способ разбиения пространства принятого сигнала на области v_1, v_2, $\ldots$, v_m.

Рассуждая аналогично, как при выборе соотношения (8.44), можно показать, что реализация условия (8.48) дает следующую систему m неравенств, обеспечивающих принятие гипотезы H_k, что $X = x_k$ (47)

$$\sum_{i=1}^{m} (r_{ij} - r_{ik}) \frac{p(x_i)\, w(Y/x_i)}{p(x_1)\, w(Y/x_1)} > 0 \quad (j = 1, 2, \ldots, m;\ j \neq k). \tag{8.49}$$

8.8. Синтез структуры решающего устройства

Как уже установлено, оптимальное решающее устройство должно строиться таким образом, чтобы оно могло вычислить функции правдоподобия $L(X)$ и отношение правдоподобия λ с последующим сравнением его с некоторым пороговым значением λ_0. Следовательно, в первую очередь решающее устройство должно вычислять условные плотности вероятности $w(Y/x_i)$. Очевидно, схема решающего устройства определяется в основном видом этой функции.

Рассмотрим общий случай многоальтернативной ситуации, когда полезный сигнал X может принимать m значений. Будем полагать помеху ξ нормальной с нулевым математическим ожиданием и аддитивной. Следовательно, принимаемый сигнал y

$$y(t) = x(t) + \xi(t).$$

Для любого отсчетного значения принятого сигнала y_i можно записать

$$y_i = x_i + \xi_i,$$

где x_i — отсчетные значения полезного сигнала; ξ_i — отсчетные значения помехи, распределенные по нормальному закону

$$w(\xi_i) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{\xi_i^2}{2\sigma_\xi^2}\right\}.$$

Вектор помехи определяется многомерным законом распределения $W(\xi_1, \xi_2, \ldots, \xi_n)$, где n — объем выборки. Полагая помеху стационарной и отсчеты некоррелированными, можно многомерный закон распределения вектора помехи представить в виде

$$w(\xi_1, \xi_2, \ldots, \xi_n) = w(\xi_1)\, w(\xi_2), \ldots, w(\xi_n) =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_\xi^2}}\right)^n \exp\left\{-\frac{\sum\limits_{i=1}^{n} \xi_i^2}{2\sigma_\xi^2}\right\}. \qquad (8.50)$$

При взаимной независимости полезного сигнала и помехи функция $w(Y/X)$ определяется (см. параграф 5.4) законом распределения помехи

$$w(Y/X) = \left(\frac{1}{\sqrt{2\pi\sigma_\xi^2}}\right)^n \exp\left\{-\frac{\sum\limits_{i=1}^{n} \xi_i^2}{2\sigma_\xi^2}\right\} = \left(\frac{1}{\sqrt{2\pi\sigma_\xi^3}}\right)^n \exp\left\{-\frac{\sum\limits_{i=1}^{n} (y_i - x_i)^2}{2\sigma_\xi^2}\right\}.$$

Для принятия оптимального решения необходимо определить отношения правдоподобия

$$\lambda_{kj} = \frac{w\,(Y/x_k)}{w\,(Y/x_j)} = \exp\left\{\frac{\sum_{i=1}^{n}(y_i - x_{ij})^2 - \sum_{i=1}^{n}(y_i - x_{ki})^2}{2\sigma_\xi^2}\right\}. \qquad (8.51)$$

Перейдем к случаю непрерывного приема сигналов в течение определенного времени T. Будем при этом полагать, что помеха типа белого шума.

Переход к непрерывному наблюдению можно осуществить с использованием теоремы Котельникова. Полагая, что отсчеты осуществляются через интервал времени $\Delta t = \frac{1}{2F_в}$, где $F_в$ — граничная частота полосы пропускания канала связи, умножаем числитель и знаменатель показателя степени e в выражении (8.51) на Δt и переходим в числителе показателя степени от суммирования отсчетов к интегрированию в пределах от 0 до T, где $T = n\Delta t = \frac{n}{2F_в}$

$$\lambda_{kj} = \exp\left\{\frac{\int_0^T [y(t) - x_j(t)]^2\,dt - \int_0^T [y(t) - x_k(t)]^2\,dt}{P_0}\right\} =$$

$$= \exp\left\{\frac{2\int_0^T y(t)\,[x_k(t) - x_j(t)]\,dt - \int_0^T [x_k(t)]^2\,dt + \int_0^T [x_j(t)]^2\,dt}{P_0}\right\} =$$

$$= \exp\left\{\frac{2\int_0^T y(t)\,[x_k(t) - x_j(t)]\,dt - E_k + E_j}{P_0}\right\}, \qquad (8.52)$$

где E_k и E_j — энергии сигналов x_k и x_j; $P_0 = \frac{\sigma_\xi^2}{F_в}$ — мощность помехи, приходящаяся на единицу полосы спектра.

Полагая энергии сигналов x_k и x_j одинаковыми, (8.52) приводим к виду

$$\lambda_{kj} = \exp\left\{\frac{2\int_0^T y(t)\,[x_k(t) - x_j(t)]\,dt}{P_0}\right\}.$$

Выбирается в качестве истинного такой сигнал x_k, для которого

$$\lambda_{kj} > \lambda_0 \ \ (j = 1,\ 2,\ \ldots,\ m;\ j \neq k). \qquad (8.53)$$

Условие (8.53) можно переписать в виде:

$$\int\limits_0^T y(t)\, x_k(t)\, dt - \int\limits_0^T y(t)\, x_j(t)\, dt > \frac{1}{2}\, p_0 \ln \lambda_0 = \lambda_0' \qquad (8.54)$$

$$(j = 1,\ 2,\ \ldots,\ m;\ j \neq k).$$

В соответствии с (8.54) структура решающего устройства должна иметь вид, как показано на рис. 8.8. Устройство содержит набор генераторов сигналов $x_1,\ x_2,\ \ldots,\ x_m$, множительных звеньев *МЗ*, интеграторов *И* и схему сравнения *СС*.

Рис. 8.8

На выходе *МЗ* получается произведение функций $y(t)\, x_i(t)$, которое затем интегрируется интегратором *И*. Схема сравнения *СС* определяет разность между различными сочетаниями выходных сигналов интеграторов, сравнивает полученные результаты со стандартным сигналом $\lambda_0' = \frac{1}{2}\, p_0 \ln \lambda_0$ и выносит решение в пользу той функции x_k, для которой выполняется условие (8.54).

Величина стандартного сигнала λ_0' определяется критерием, положенным в основу синтеза решающего устройства. В частности, если в качестве такового используется критерий идеального наблюдателя (критерий Котельникова), то

$$\lambda_0' = \frac{1}{2}\, p_0 \ln \frac{p(x_0)}{p(x_1)}. \qquad (8.55)$$

Решающее устройство, реализуемое на базе критерия идеального наблюдателя, обеспечивает минимальную вероятность ошибки принятия решения. Такие устройства принято называть идеальными приемниками.

8.9. Восстановление сигналов

Восстановление сигналов сводится к оценке некоторого числа неизвестных параметров полезного сигнала. Ограничимся рассмотрением случая оценки одного из параметров сигнала, например амплитуды B, при заданной форме сигнала. При этом помехи будем

полагать аддитивными типа белого гауссова шума. Представим полезный сигнал в виде

$$x(t) = Bf(t),$$

где $f(t)$ — известная функция времени; B — параметр сигнала.

Задача состоит в том, чтобы по принятой выборке Y определить, каково значение параметра B в полезном сигнале X.

В отличие от случаев обнаружения и различия сигналов здесь имеет место бесконечное множество возможных значений параметра B и, соответственно, бесконечное множество гипотез. Методы, рассматриваемые в случае двухальтернативных и многоальтернативных ситуаций, применимы и для задачи восстановления сигнала.

Произведем оценку параметра B методом максимума правдоподобия. Если отсчет принятого сигнала производится в дискретные моменты времени, то функция правдоподобия для параметра B будет равна

$$L(B) = w(Y/B) = \left(\frac{1}{\sqrt{2\pi\sigma_\xi^2}}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}[y(t_i) - Bf(t_i)]^2}{2\sigma_\xi^2}\right\}. \quad (8.56)$$

Задача состоит в том, чтобы найти такое значение параметра B для которого функция правдоподобия максимальна. Максимуму функции правдоподобия соответствует минимальное значение показателя степени в выражении (8.56)

$$\Phi(B) = \frac{1}{2\sigma_\xi^2}\sum_{i=1}^{n}[y(t_i) - Bf(t_i)]^2 = \frac{1}{2\sigma_\xi^2}\left\{\sum_{i=1}^{n}[y(t_i)]^2 - 2B\sum_{i=1}^{n} y(t_i) f(t_i) + B^2\sum_{i=1}^{n}[f(t_i)]^2\right\} = \text{мин.}$$

Из условия минимума

$$\frac{\partial\Phi(B)}{\partial B} = 0$$

имеем

$$-2\sum_{i=1}^{n} y(t_i) f(t_i) + 2B\sum_{i=1}^{n}[f(t_i)]^2 = 0,$$

откуда получаем оценочное значение параметра

$$B^* = \frac{\sum_{i=1}^{n} y(t_i) f(t_i)}{\sum_{i=1}^{n}[f(t_i)]^2}. \quad (8.57)$$

Осуществив переход к непрерывному примеру, получим

$$B^* = \frac{\int\limits_0^T y(t) f(t)\, dt}{\int\limits_0^T [f(t)]^2\, dt} = \frac{\int\limits_0^T y(t) f(t)\, dt}{E_0}, \tag{8.58}$$

где $E_0 = \int\limits_0^T [f(t)]^2\, dt$ — удельная энергия сигнала (энергия сигнала при амплитуде $B = 1$).

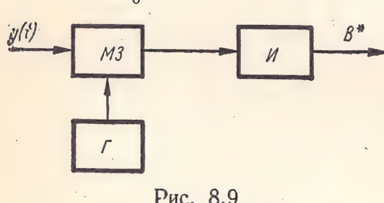


Рис. 8.9

На рис. 8.9 приведена схема решающего устройства, осуществляющего операцию оценки параметра сигнала. Устройство содержит генератор сигнала $f(t)$, множительное звено *МЗ*, осуществляющее умножение $y(t)$ на $f(t)$, и интегратор, производящий интегрирование произведения $y(t) f(t)$.

Для оценки точности восстановления сигнала используем критерий среднеквадратического отклонения. С этой целью в (8.58) принимаемый сигнал выразим в виде суммы $y(t) = Bf(t) + \xi(t)$. Тогда

$$B^* = \frac{B}{E_0} \int\limits_0^T [f(t)]^2\, dt + \frac{1}{E_0} \int\limits_0^T \xi(t) f(t)\, dt = B + \frac{1}{E_0} \int\limits_0^T \xi(t) f(t)\, dt.$$

Погрешность восстановления

$$B^* - B = \frac{1}{E_0} \int\limits_0^T \xi(t) f(t)\, dt.$$

Дисперсия погрешности

$$\sigma^2 = m_1 \left\{ \frac{1}{E_0^2} \left[\int\limits_0^T \xi(t) f(t)\, dt \right]^2 \right\} = m_1 \left\{ \frac{1}{E_0^2} \int\limits_0^T \xi(t_1) f(t_1)\, dt_1 \times \right.$$

$$\left. \times \int\limits_0^T \xi(t_2) f(t_2)\, dt_2 \right\} = \frac{1}{E_0^2} \int\limits_0^T \int\limits_0^T m_1 \{\xi(t_1) \xi(t_2)\} f(t_1) f(t_2)\, dt_1 dt_2.$$

Среднее от произведения $\xi(t_1)\xi(t_2)$ представляет корреляционную функцию помехи

$$m_1 \{\xi(t_1) \xi(t_2)\} = K\xi(t) = G_0 \delta(\tau),$$

где G_0 — спектральная плотность помехи; $\delta(\tau)$ — дельта-функция;

$$\tau = t_2 - t_1.$$

Тогда

$$\sigma^2 = \frac{G_0}{E_0^2} \int\limits_0^T \int\limits_0^T \delta(t_2 - t_1) f(t_1) f(t_2)\, dt_1 dt_2 =$$

$$= \frac{G_0}{E_0^2} \int\limits_0^T f(t_1) \left[\int\limits_0^T \delta(t_2 - t_1) f(t_2)\, dt_2 \right] dt_1 = \frac{G_0}{E_0^2} \int\limits_0^T [f(t_1)]^2\, dt = \frac{G_0}{E_0}.$$

Следовательно, среднеквадратическое значение погрешности восстановления

$$\sigma = \sqrt{\frac{G_0}{E_0}}. \qquad (8.59)$$

Задача восстановления сигнала может быть также решена методом оптимальной фильтрации. В параграфе 8.5. рассматривались оптимальные фильтры, обеспечивающие при известной форме сигнала максимально возможное отношение амплитудного значения сигнала к действующему значению помехи. Такие фильтры получили наименование согласованных (относительно формы сигнала), или фильтров Норса. Фильтры этого типа целесообразно использовать при решении задач различения или обнаружения сигналов.

Максимизируя отношение сигнал /помеха даже ценой существенного искажения формы сигнала, эти фильтры обеспечивают уменьшение ошибок при различении или обнаружении сигналов.

Однако при решении задачи восстановления важное значение имеет обеспечение минимальной ошибки оценки параметров полезного сигнала.

Эта задача решается с помощью фильтров Колмогорова — Винера. Критерием оптимальности данных фильтров является получение минимальной среднеквадратической ошибки воспроизведения сигналов. Эти фильтры не только очищают сигнал от помехи, но и предсказывают его значение на некоторое время вперед. В связи с этим их часто называют фильтрами сглаживания и предсказания. При реализации таких фильтров считается, что полезный сигнал является случайным процессом с известными вероятностными характеристиками.

Показано [37, 48], что если на вход фильтра поступает сумма сигнала и помехи, являющихся стационарными случайными функциями времени, то минимальная среднеквадратическая ошибка воспроизведения сигнала будет при условии, если функция передачи фильтра выражается соотношением

$$K(j\omega) = \frac{S_x(j\omega)}{S_x(j\omega) + S_\xi(j\omega)} \exp\{j\omega T\}, \qquad (8.60)$$

где $S_x(j\omega)$ и $S_\xi(j\omega)$ — спектральные плотности полезного сигнала и помехи; T — время предсказания.

При этом среднеквадратическая ошибка воспроизведения сигнала

$$\sigma = \sqrt{\int_0^\infty \frac{S_x(j\omega)\, S_\xi(j\omega)}{S_x(j\omega) + S_\xi(j\omega)}\, d\omega}\,. \qquad (8.61)$$

Отсюда видно, что идеальная фильтрация ($\sigma = 0$) возможна только в том случае, когда

$$S_x(j\omega) \cdot S_\xi(j\omega) = 0,$$

т. е. когда спектры сигнала и помехи не перекрываются.

8.10. Примеры

Пример 8.1. По каналу связи, в котором действует аддитивная стационарная помеха, передается периодическая последовательность прямоугольных импульсов. Параметры полезного сигнала: величина $A_x = 2$В, период следования $T_n = 100$ мс. Помеха имеет нормальный закон распределения. Среднеквадратическое значение помехи $\sigma_\xi = 5$В, математическое ожидание $m_\xi = 0$. Обработка сигналов на приемной стороне осуществляется методом синхронного накопления.

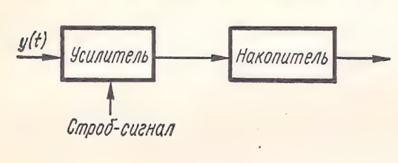


Рис. 8.10

Определить время обработки сигналов, необходимое для обеспечения превышения сигнала над помехой в 4 раза.

Решение. Упрощенная схема приемника с синхронным накоплением представлена на рис. 8.10. Усилитель приемника в исходном состоянии заперт. Строб-сигнал поступает синхронно с полезным сигналом и обеспечивает отпирание приемника на время подачи полезного сигнала.

Отношение сигнал/помеха на выходе накопителя в случае стационарной помехи

$$\left(\frac{A_x}{A_\xi}\right)_{\text{вых}} = \sqrt{n}\sqrt{\left(\frac{P_x}{P_\xi}\right)_{\text{вх}}}\,,$$

где $\left(\frac{P_x}{P_\xi}\right)_{\text{вх}}$ — отношение мощностей сигнала и помехи на входе приемника; n — количество отсчетов за время приема.

В нашем случае

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вх}} = \left(\frac{A_x}{\sigma_\xi}\right)^2 = \left(\frac{2}{5}\right)^2.$$

Продолжительность обработки сигналов в приемнике

$$T_{пр} = nT_n = \frac{\left(\frac{A_x}{A_\xi}\right)^2_{вых}}{\left(\frac{A_x}{\sigma_\xi}\right)^2_{вх}} T_n.$$

По условию задачи

$$\left(\frac{A_x}{A_\xi}\right)_{вых} = 4.$$

Тогда необходимое время наблюдения

$$T_{пр} = \frac{4^2}{\left(\frac{2}{5}\right)^2} \cdot 0{,}1 = 10 \text{ с}.$$

Пример 8.2. Необходимо осуществить обнаружение постоянного сигнала величиной $a = 2$В на фоне аддитивной помехи с нормальным распределением и средним значением, равным нулю. Метод приема — однократный отсчет. Произвести синтез приемного устройства, работающего на основе критерия максимума правдоподобия, и определить пороговый уровень.

Решение. Так как по условию задачи помеха аддитивна и выборка Y представляет одномерную величину, то функции правдоподобия $L(a)$ и $L(0)$ определяются законом распределения помехи

$$L(a) = w(Y/a) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{(Y-a)^2}{2\sigma_\xi^2}\right\}; \tag{8.62}$$

$$L(0) = w(Y/0) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{Y^2}{2\sigma_\xi^2}\right\}. \tag{8.63}$$

Отношение правдоподобия при этом

$$\lambda = \frac{L(a)}{L(0)} = \exp\left\{-\frac{(Y-a)^2 - Y^2}{2\sigma_\xi^2}\right\} = \exp\left\{\gamma\left[\frac{Y}{a} - \frac{1}{2}\right]\right\}, \tag{8.64}$$

где $\gamma = \frac{a^2}{\sigma_\xi^2}$.

Графики функций (8.62) — (8.64) приведены на рис. 8.11.

При использовании критерия максимума правдоподобия пороговое значение отношения правдоподобия $\lambda_0 = 1$. Тогда, приравняв (8.64) единице, получим условие для порогового значения входного сигнала $Y_{пор}$:

$$\exp\left\{\gamma\left[\frac{Y_{пор}}{a} - \frac{1}{2}\right]\right\} = 1,$$

откуда получаем

$$\frac{Y_{пор}}{a}-\frac{1}{2}=0,$$

или

$$Y_{пор}=\frac{a}{2}=1В.$$

Таким образом, приемное устройство представляет собой устройство сравнения, сравнивающее входной сигнал с пороговым уровнем $Y_{пор}$, равным половине величины полезного сигнала.

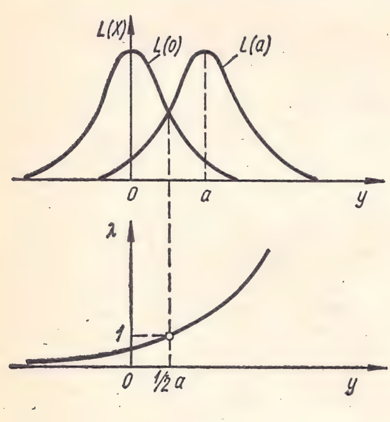


Рис. 8.11

Если входной сигнал превышает пороговый уровень в 1В, то выносится решение, что во входном сигнале имеется полезный сигнал. Если же входной сигнал равен или меньше $Y_{пор}$, то выносится решение об отсутствии полезного сигнала.

Пример 8.3. Решить предыдущую задачу для случая, когда используется критерий идеального наблюдателя и дополнительно известно, что отношение априорных вероятностей отсутствия и присутствия полезного сигнала в принятом сигнале $\frac{p(0)}{p(a)}=10$ и среднеквадратическое значение помех ξ равно $\sigma_\xi=0{,}5$ В.

Решение. При использовании критерия идеального наблюдателя пороговое значение отношения правдоподобия $\lambda_0=\frac{p(0)}{p(a)}$.

Тогда, приравняв правую часть (8.64) пороговому значению λ_0, получим следующее выражение для порогового уровня

$$Y_{пор}=a\left(\frac{1}{\gamma}\ln\lambda_0+\frac{1}{2}\right)=2\left[\left(\frac{0{,}5}{2}\right)^2\ln 10+\frac{1}{2}\right]=1{,}88\,В.$$

Пример 8.4. Решить задачу, приведенную в примере 8.2, для случая, когда используется критерий Неймана—Пирсона и дополнительно известно, что среднеквадратическое значение помех $\sigma_\xi=1{,}25$ В и вероятность ошибки первого рода принятия решений не должна превышать величины $\varepsilon=0{,}05$.

Решение. Искомое пороговое значение $Y_{пор}$ можно определить из соотношения (8.39), если перейти от переменных λ к переменным Y:

$$\alpha=\int\limits_{Y_{пор}}^{\infty} w(Y/0)\,dY=\varepsilon.$$

Подставляя из (8.63) выражение для плотности $w(Y/0)$, получим

$$\alpha = \int_{Y_{\text{пор}}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{Y^2}{2\sigma_\xi^2}\right\} dY = \varepsilon.$$

После замены $\frac{Y}{\sqrt{\sigma_\xi^2}} = z$ получаем уравнение для нахождения $Y_{\text{пор}}$:

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{Y_{\text{пор}}}{\sigma_\xi}}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{1}{2}\left[1 - 2\Phi\left(\frac{Y_{\text{пор}}}{\sigma_\xi}\right)\right] = \varepsilon,$$

где $\Phi\left(\frac{Y_{\text{пор}}}{\sigma_\xi}\right)$ — функция Лапласа.

По условию задачи

$$\Phi\left(\frac{Y_{\text{пор}}}{\sigma_\xi}\right) = \frac{1}{2} - \varepsilon = 0{,}45.$$

По таблицам функций Лапласа находим

$$\frac{Y_{\text{пор}}}{\sigma_\xi} = 1{,}65,$$

откуда искомый пороговый уровень

$$Y_{\text{пор}} = 1{,}65\,\sigma_\xi = 1{,}65 \cdot 1{,}25 = 2{,}06 \text{ В}.$$

Пример 8.5. Произвести синтез приемного устройства, обеспечивающего обнаружение детерминированного сигнала по критерию идеального наблюдателя. Определить необходимую величину порогового уровня, ошибки первого и второго рода и общую ошибку принятия решения, если задано следующее:

метод приема — непрерывный в течение времени $T_{\text{пр}} = 1$С;

величина полезного сигнала $A = 8$В;

априорная вероятность присутствия сигнала $p(A) = 0{,}2$;

помеха аддитивная типа белого гауссова шума с удельной мощностью $p_0 = 50\,\frac{\text{В}^2}{\text{Гц}}$ и нулевым средним значением.

Решение. Отношение правдоподобия для случая двухальтернативной ситуации и непрерывного приема имеет вид

$$\lambda = \exp\left\{\frac{\int_0^{T_{\text{пр}}} [y(t)]^2\,dt - \int_0^{T_{\text{пр}}} [y(t) - x(t)]^2\,dt}{P_0}\right\} =$$

$$= \exp\left\{\frac{2\int\limits_0^{T_{\text{пр}}} y(t)\,x(t)\,dt - \int\limits_0^{T_{\text{пр}}} [x(t)]^2\,dt}{p_0}\right\} = \exp\left\{\frac{2\int\limits_0^{T_{\text{пр}}} y(t)\,x(t)\,dt - E_x}{p_0}\right\},$$

где E_x — энергия полезного сигнала.

Принимается решение о наличии сигнала, если

$$\lambda > \lambda_0 = \frac{p(0)}{p(A)},$$

или

$$\frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} y(t)\,x(t)\,dt > \ln\frac{p(0)}{p(A)} + \frac{E_x}{p_0} = \lambda_0'. \qquad (8.65)$$

В соответствии с (8.65), приемное устройство должно содержать: генератор сигнала $x(t)$, множительное звено *МЗ*, интегратор *И* и устройство сравнения *СС*.

Так как генератор *Г* создает постоянный сигнал $x(t) = A$, то в условиях данной задачи выражение (8.65) приобретает вид

$$\int\limits_0^{T_{\text{пр}}} y(t)\,dt > \frac{p_0}{2A}\ln\frac{p(0)}{p(A)} + \frac{AT_{\text{пр}}}{2} = \lambda_0'.$$

Следовательно, при $x(t) = A = \text{const}$ схема приемника упрощается и приобретает вид, изображенный на рис. 8.12.

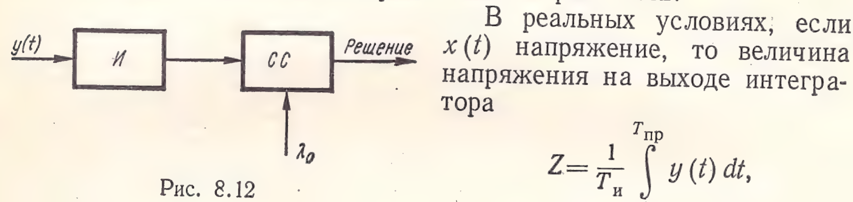


Рис. 8.12

В реальных условиях, если $x(t)$ напряжение, то величина напряжения на выходе интегратора

$$Z = \frac{1}{T_{\text{и}}}\int\limits_0^{T_{\text{пр}}} y(t)\,dt,$$

где $T_{\text{и}}$ — постоянная времени интегратора.

Тогда пороговый сигнал должен быть соответственно

$$\lambda_0' = \frac{p_0}{2AT_{\text{и}}}\ln\frac{p(0)}{p(A)} + \frac{AT_{\text{пр}}}{2T_{\text{и}}}.$$

Принимая постоянную времени интегратора $T_{\text{и}} = 10\text{с}$, получим

$$\lambda_0' = \frac{50}{2\cdot 8\cdot 10}\ln 4 + \frac{8}{2\cdot 10} = 0{,}8\text{В}.$$

Условная вероятность ошибки первого рода — это вероятность того, что при отсутствии полезного сигнала справедливо неравенство (8.65) при $y(t) = \xi(t)$, где $\xi(t)$ — помеха.

Интеграл $\int\limits_0^{T_{\text{пр}}} \xi(t)\,x(t)\,dt$ получен из суммы, у которой все слагаемые подчинены нормальному закону распределения с нулевым средним. Тогда сумма, а следовательно, и величина $\eta_1 = \frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} \xi(t)x(t)\,dt$ подчинены нормальному закону распределения с нулевым средним.

В соответствии со значением дисперсии интеграла $\int\limits_0^{T_{\text{пр}}} y(t)\,x(t)\,dt$, полученым в параграфе 8.11, дисперсия величины η_1

$$\sigma^2_{\eta_1} = \frac{2E_x}{p_0}.$$

Таким образом, выражение для условной вероятности ошибки первого рода можно представить в виде

$$\alpha = \int\limits_{\lambda_0'}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2_{\eta_1}}} \exp\left\{-\frac{\eta_1^2}{2\sigma^2_{\eta_1}}\right\} d\eta_1. \qquad (8.66)$$

Условная вероятность ошибки второго рода — это вероятность того, что при наличии полезного сигнала справедливо неравенство

$$\frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} y(t)\,x(t)\,dt \leqslant \ln\frac{p(0)}{p(A)} + \frac{E_x}{p_0} = \lambda_0',$$

где

$$y(t) = x(t) + \xi(t).$$

Рассмотрим случайную величину

$$\eta_2 = \frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} [x(t) + \xi(t)]\,x(t)\,dt = \frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} [x(t)]^2\,dt + \frac{2}{p_0} \times$$

$$\times \int\limits_0^{T_{\text{пр}}} \xi(t)\,x(t)\,dt = \frac{2E_x}{p_0} + \frac{2}{p_0}\int\limits_0^{T_{\text{пр}}} \xi(t)\,x(t)\,dt = \frac{2E_x}{p_0} + \eta_1.$$

Таким образом, случайная величина η_2 содержит постоянную составляющую, равную $\frac{2E_x}{p_0}$, и переменную составляющую, распределенную по нормальному закону с дисперсией, равной $\frac{2E_x}{p_0}$. Следовательно, можно утверждать, что η_2 — случайная величина рас-

пределенная по нормальному закону с математическим ожиданием, равным $\frac{2E_x}{p_0}$, и дисперсией

$$\sigma_{\eta_2}^2 = \frac{2E_x}{p_0}.$$

Тогда условная вероятность ошибки второго рода может быть представлена в виде

$$\beta = \int_{-\infty}^{\lambda_0'} \frac{1}{\sqrt{2\pi\sigma_{\eta_2}^2}} \exp\left\{-\frac{\left(\eta_2 - \frac{2E_x}{p_0}\right)^2}{2\sigma_{\eta_2}^2}\right\} d\eta_2. \qquad (8.67)$$

Переходя к новым переменным $z_1 = \frac{\eta_1}{\sigma_{\eta_1}}$ и $z_2 = \frac{\eta_2 - \frac{2E_x}{p_0}}{\sigma_{\eta_2}}$, получим следующие выражения для ошибок первого и второго рода:

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda_0'}{\sigma_{\eta_1}}}^{\infty} \exp\left\{-\frac{z_1^2}{2}\right\} dz_1 = \frac{1}{2} - \Phi\left[\frac{\ln\frac{p(0)}{p(A)} + \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right]; \qquad (8.68)$$

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\lambda_0' - \frac{2E_x}{p_0}}{\sigma_{\eta_2}^2}} \exp\left\{-\frac{z_2^2}{2}\right\} dz_2 = \frac{1}{2} + \Phi\left[\frac{\ln\frac{p_0}{p(A)} - \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right]. \qquad (8.69)$$

Функции Лапласа, входящие в выражения (8.68) и (8.69), определяем по таблицам, приведенным в приложении 1.

$$\Phi\left[\frac{\ln\frac{p(0)}{p(A)} + \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right] = \Phi(1{,}72) = 0{,}457;$$

$$\Phi\left[\frac{\ln\frac{p(0)}{p(A)} - \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right] = \Phi(0{,}06) = 0{,}024.$$

Условные вероятности ошибок первого и второго рода

$$\alpha = \frac{1}{2} - 0{,}457 = 0{,}043; \qquad \beta = \frac{1}{2} + 0{,}024 = 0{,}524.$$

Общая ошибка принятия решения

$$p_{\text{ош}} = p(0)\,\alpha + p(A)\,\beta = 0{,}8 \cdot 0{,}043 + 0{,}2 \cdot 0{,}524 = 0{,}139.$$

Пример 8.6 [10]. На вход приемного устройства поступает смесь полезного гармонического сигнала $x(t) = A\sin\omega_0 t$ с амплитудой $A = 0{,}4$В и аддитивной помехи, распределенной по нормальному закону с нулевым средним и среднеквадратическим значением $\sigma_\xi = 0{,}2$В. Произведено два замера входного сигнала $y(t) = x(t) + \xi(t)$ в моменты времени $t_1 = \frac{T_0}{4}$ и $t_2 = \frac{3T_0}{4}$ (рис. 8.13), где $T_0 = \frac{2\pi}{\omega_0}$ — период сигнала $x(t)$. Априорные вероятности значений $x_1 = 0$ и $x_2 = A = 0{,}4$В соответственно $p(x_1) = p(x_2) = 0{,}5$.

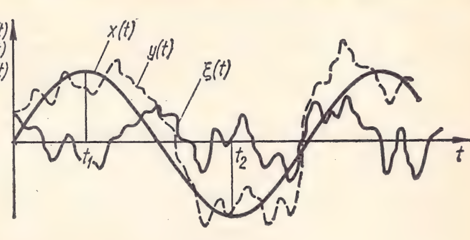


Рис. 8.13

Найти апостериорные вероятности значений $x_1 = 0$ и $x_2 = A$ после указанных выше замеров.

Решение. Искомые апостериорные вероятности значений могут быть определены по формуле Байеса (см. выражение (8.18))

$$p(x_1/Y) = \frac{p(x_1)\, w(Y/x_1)}{p(x_1)\, w(Y/x_1) + p(x_2)\, w(Y/x_2)};$$

$$p(x_2/Y) = \frac{p(x_2)\, w(Y/x_2)}{p(x_1)\, w(Y/x_1) + p(x_2)\, w(Y/x_2)}.$$

Условная плотность вероятности выборки $y_1 = y(t_1)$, $y_2 = y(t_2)$ при $x = x_1 = 0$ и при $x = x_2 = A$ в предположении, что отсчеты помех $\xi(t_1)$ и $\xi(t_2)$ статистически независимы, соответственно

$$w\left(\frac{y_1 y_2}{x_1}\right) = w\left(\frac{y_1}{x_1}\right) w\left(\frac{y_2}{x_1}\right) = \frac{1}{\left(\sqrt{2\pi\sigma_\xi^2}\right)^2} \exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\};$$

$$w\left(\frac{y_1 y_2}{x_2}\right) = w\left(\frac{y_1}{x_2}\right) w\left(\frac{y_2}{x_2}\right) = \frac{1}{\left(\sqrt{2\pi\sigma_\xi^2}\right)^2} \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}.$$

Следовательно, искомые апостериорные вероятности

$$P\left(\frac{0}{Y}\right) = \frac{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\}}{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\} + \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}} =$$

$$= \frac{\exp(-1{,}63)}{\exp(-1{,}63) + \exp(-0{,}626)} = 0{,}268;$$

$$P\left(\frac{A}{Y}\right) = \frac{\exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}}{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\} + \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}} =$$

$$= \frac{\exp(-0{,}626)}{\exp(-1{,}63) + \exp(-0{,}626)} = 0{,}732.$$

Пример 8.7 [10]. В приемнике, предназначенном для обнаружения сигнала и работающем по критерию Неймана—Пирсона, установлен порог λ_0', соответствующий значению вероятности ошибки первого рода $\alpha = 0{,}05$. Априорные вероятности значений $x_1 = 0$ и $x_2 = A$ равны $p(0) = p(A) = 0{,}5$. Величина $\rho = \frac{E_x}{p_0} = 5$. Определить вероятность ошибки второго рода и общей ошибки принятия решения.

Решение. Из формулы (8.68) находим значение функции Лапласа

$$\Phi\left[\frac{\ln\frac{p(0)}{p(A)} + \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right] = \frac{1}{2} - \alpha = 0{,}45.$$

Значение аргумента этой функции (см. приложение 1) равно 1,645. Тогда

$$\ln\frac{p(0)}{p(A)} = \sqrt{2\frac{E_x}{p_0}} \cdot 1{,}645 - \frac{E_x}{p_0} = \sqrt{10} \cdot 1{,}645 - 5 = 0{,}198.$$

Условная вероятность ошибки второго рода

$$\beta = \frac{1}{2} + \Phi\left[\frac{\ln\frac{p(0)}{p(A)} - \frac{E_x}{p_0}}{\sqrt{2\frac{E_x}{p_0}}}\right] = 1 + \Phi\left(\frac{0{,}198 - 5}{\sqrt{10}}\right) = 0{,}564.$$

Общая безусловная вероятность ошибки

$$p_{\text{ош}} = p(0)\,\alpha + p(A)\,\beta = 0{,}5\,(0{,}05 + 0{,}564) = 0{,}307.$$

Пример 8.8 [10]. Производится оценка амплитуды A сигнала $x(t) = A\sin\omega_0 t$. Время наблюдения $T_{\text{пр}} \gg T_0 = \frac{2\pi}{\omega_0}$. Найти каково должно быть среднее значение параметра $A_{\text{ср}}$, чтобы относительная погрешность оценки A

$$\gamma_\sigma = \frac{\sigma}{A_{\text{ср}}} = 0{,}01.$$

Решение. При $T_{\text{пр}} \gg T_0$, $T_{\text{пр}} = nT_0$, где $n \gg 1$.

Тогда энергия сигнала

$$E_x = n\int_0^{T_0}[A\sin\omega_0 t]^2\,dt = A^2 n\int_0^{T_1}\sin^2\omega_0 t\,dt =$$

$$= A^2 n\int_0^{T_0}\left(\frac{1}{2}-\frac{1}{2}\cos 2\omega_0 t\right)dt = A^2 n\,\frac{1}{2}\,T_0 = \frac{A^2 T_{\text{пр}}}{2}.$$

Абсолютная погрешность оценки параметра определяется по формуле (8.59)

$$\sigma = \sqrt{\frac{G_{10}}{E_0}},$$

где

$$E_0 = \frac{E_x}{A^2} = \frac{T_{\text{пр}}}{2}.$$

Тогда

$$\sigma = \sqrt{\frac{2G_{10}}{T_{\text{пр}}}} = \sqrt{\frac{2-0{,}2}{2}} = 0{,}447 \text{ мВ}.$$

Следовательно

$$A_{\text{ср}} = \frac{\sigma}{\gamma_\sigma} = \frac{0{,}447}{0{,}01} \cong 0{,}45 \text{ мВ}.$$

Пример 8.9 [27]. На вход оптимального фильтра Норса поступает сигнал $X(t)$ (рис. 8.14)

$$x(t) = \begin{cases} \dfrac{1}{RC}\exp\left\{-\dfrac{t-T}{RC}\right\} & \text{при } t \leqslant T; \\ 0 & \text{при } t > T. \end{cases}$$

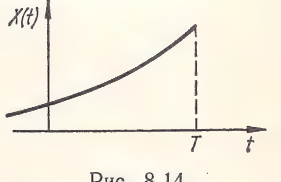


Рис. 8.14

Необходимо определить передаточную функцию фильтра и сигнал на его выходе при условии, что в канале действует белый гауссов шум.

Решение. Передаточную функцию фильтра определяем с помощью выражения (8.11).

Спектральная плотность сигнала определяется по формуле преобразования Фурье

$$\dot{S}_x(j\omega) = \int_{-\infty}^{\infty} x(t)\exp(-j\omega t)\,dt = \frac{\exp\left(-\dfrac{T}{RC}\right)}{RC}\times$$

$$\times\int_{-\infty}^{\infty}\exp\left[\left(\frac{1}{RC}-j\omega\right)t\right]dt = \frac{\exp(-j\omega T)}{1-j\omega RC}.$$

Комплексно-сопряженная спектральная плотность сигнала

$$S_x^*(j\omega)=\frac{\exp(j\omega T)}{1+j\omega RC}.$$

С точностью до постоянного множителя b передаточная функция фильтра

$$K(j\omega)=bS_x^*(j\omega)\exp\{-j\omega T\}=\frac{b}{1+j\omega RC}.$$

Такой передаточной функции соответствует схема обычной интегрирующей цепочки RC, вид которой изображен на рис. 8.15.

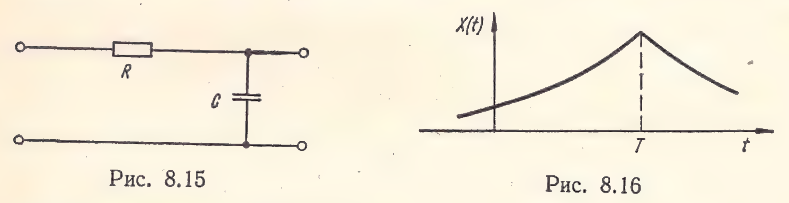


Рис. 8.15 Рис. 8.16

Сигнал на выходе фильтра

$$x(t)_{\text{вых}}=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_x(j\omega)K(j\omega)\exp\{-j\omega T\}\,d\omega$$

и аналитически описывается функцией

$$x(t)_{\text{вых}}=\begin{cases}\frac{1}{2RC}\exp\left\{-\frac{T-t}{RC}\right\} \text{ при } t\leqslant T;\\ \frac{1}{2RC}\exp\left\{-\frac{t-T}{RC}\right\} \text{ при } t>T.\end{cases}$$

Графическое изображение выходного сигнала дано на рис. 8.16.

Пример 8.10 [27]. На вход оптимального фильтра Норса поступает сигнал прямоугольной формы длительности τ_0 с амплиту-

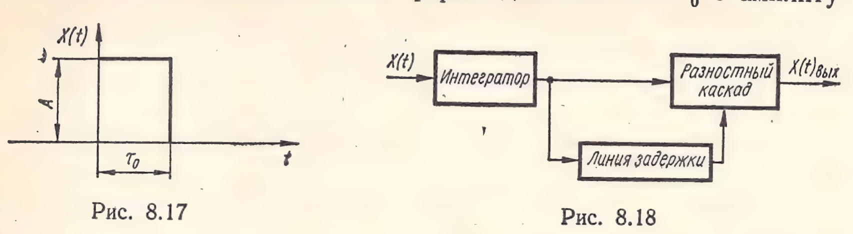


Рис. 8.17 Рис. 8.18

дой A (рис. 8.17). Необходимо определить передаточную функцию фильтра при условии, что в канале действует белый гауссов шум.

Решение. Спектральная плотность сигнала

$$S_x(j\omega)=\int_{-\infty}^{\infty}x(t)\exp\{-j\omega t\}\,dt=A\int_{0}^{\tau_0}\exp\{-j\omega t\}\,dt=$$

$$=-\frac{A}{j\omega}(1-\exp\{-j\omega\tau_0\}).$$

Комплексно-сопряженная спектральная плотность

$$S_x^*(j\omega) = \frac{A}{j\omega}(\exp\{j\omega\tau_0\} - 1).$$

Передаточная функция фильтра

$$K(j\omega) = bS_x^*(j\omega)\exp\{-j\omega T\} = \frac{bA}{j\omega}(1 - \exp\{-j\omega\tau_0\}) =$$
$$= \frac{B}{j\omega}(1 - \exp\{-j\omega\tau_0\}),$$

где $B = bA$.

Такая передаточная функция реализуется схемой, состоящей из интегрирующей ячейки с коэффициентом передачи, равным $\frac{B}{j\omega}$, линии задержки, задерживающей сигнал на время τ_0, и разностного каскада (рис. 8.18).

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. В чем заключается сущность частотной фильтрации сигналов?
2. Как связано улучшение отношения сигнал/помеха при частотной фильтрации со временем обработки сигналов?
3. В чем заключается сущность метода накопления и в каких случаях целесообразно его применять?
4. Из каких соображений нужно выбирать интервалы между отсчетами при накопительном приеме?
5. Чем отличается метод интегрального приема от метода накопления и в каких случаях он целесообразен?
6. За счет чего достигается увеличение отношения сигнал/помеха при применении методов накопления и интегрального приема?
7. В чем заключается сущность корреляционного метода приема и в каких случаях целесообразно его применять?
8. Из каких соображений выбирается время сдвига при корреляционном методе приема?
9. Какие фильтры называются согласованными?
10. Для выделения каких сигналов применяются фильтры Норса?
11. Как связана переходная функция фильтра Норса с входным сигналом?
12. В чем сущность задачи проверки гипотез?
13. Что понимается под пространством наблюдения, пространством полезного сигнала и пространством помехи?
14. Что понимается под выборкой и объемом выборки?
15. Что такое функция правдоподобия и отношение правдоподобия?
16. Что такое ошибки первого и второго рода и как они количественно оцениваются?
17. Как количественно оценивается полная ошибка принятия решения?
18. Как зависят ошибки первого и второго рода от порогового уровня?
19. В чем состоит сущность задачи обнаружения сигнала?
20. В чем заключается сущность критерия максимума правдоподобия и каковы его достоинства?
21. В чем заключается сущность критерия максимума апостериорной вероятности и каковы его преимущества перед критерием максимума правдоподобия?
22. В чем заключается сущность критерия идеального наблюдателя?
23. В чем заключается отличие критерия идеального наблюдателя от критерия максимума апостериорной вероятности; что общего у этих критериев?
24. В чем заключается сущность критерия Неймана — Пирсона и в каких случаях целесообразно этот критерий применять?

25. Что понимается под риском?
26. В чем заключается сущность критерия минимального риска?
27. В чем заключается сущность минимаксного критерия?
28. В чем состоит сущность задачи различения сигналов?
29. Что понимается под идеальным приемником?
30. В чем заключается сущность задачи восстановления сигналов?
31. Что понимается под фильтрами Колмогорова — Винера?
32. В каких случаях возможна идеальная фильтрация с помощью фильтров Колмогорова — Винера?

Глава IX

ИНФОРМАЦИОННАЯ ОЦЕНКА АВТОМАТИЗИРОВАННЫХ СИСТЕМ КОНТРОЛЯ И УПРАВЛЕНИЯ

9.1. Общая характеристика автоматизированных систем контроля и управления (АСКУ)

В последнее время для получения информации о положении и состоянии объектов в пространстве и времени все более широкое применение находят сложные системы контроля и управления. Отдельные сравнительно простые автоматы на базе вычислительной техники объединяются в большие автоматизированные системы контроля и управления [16].

К таким системам в первую очередь следует отнести автоматические системы стабилизации летательных аппаратов, технологические и организационно-экономические автоматизированные системы управления нефтяными районами, энергосетью страны, цехами, предприятиями, фирмами, отраслями народного хозяйства, страной в целом.

Все эти системы не только имеют множество сложных устройств и органов, но и высокий уровень организации, сложные функциональные взаимосвязи устройств и органов.

Для того чтобы правильно судить о ходе процесса контроля и управления, а также о состоянии и положении контролируемого и управляемого объекта, необходимо контролировать большое число параметров различной физической природы. Технические процессы контроля и управления сложными объектами тесно связаны между собой и не всегда контроль и управление ими можно поручить нескольким действующим независимо операторам и автоматам. Поэтому большое число измерительных и управляющих приборов приходится сосредоточивать в одном месте. Применение централизованных автоматизированных систем контроля и управления избавляет от необходимости пользования показаниями множества приборов и транспарантов.

Основное назначение АСКУ — правильно определить в течение заданного интервала времени состояние и положение объекта

и управлять им с учетом состояния. В случае появления отклонений или неисправностей АСКУ должна обеспечивать обнаружение отклонения или отказа места неисправности с заданной точностью.

Все чаще в последнее время АСКУ используются также для прогнозирования работоспособности и положения объекта в целом или его функциональных элементов. В этих случаях она должна предсказывать поведение параметров или объекта в целом на заданное время вперед.

АСКУ позволит кроме эффективного контроля и управления вести оптимизацию этих процессов, а также накапливать информацию для прогнозирования отказов системы контроля и управления объектом.

Процесс контроля и управления сложными объектами можно рассматривать как процесс формирования и регулирования физически независимых или взаимосвязанных процессов, происходящих одновременно либо с некоторым сдвигом во времени. Техническую систему контроля и управления сложными объектами необходимо в этом случае рассматривать как многомерную систему.

По целевому назначению АСКУ можно разделять на системы: стабилизации, программного управления, следящие, оптимальные и смешанные.

Многомерные АСКУ стабилизации предназначены для одновременного поддержания с определенной точностью многих выходных параметров. Многомерные АСКУ программного управления предназначены для автоматического управления несколькими физическими процессами по заданной программе с определенной точностью. Многомерные следящие АСКУ предназначены для одновременного воспроизведения всеми выходными параметрами неизвестных заранее различных законов управления. Оптимальные многомерные АСКУ имеют оптимизатор, позволяющий получать оптимальные значения выходных параметров. Многомерные смешанные АСКУ могут одновременно следить, стабилизировать, управлять по программе и оптимизировать множество параметров.

Многомерные АСКУ — это сложные информационные системы, характеризуемые интенсивными потоками информации.

В связи с этим представляет интерес произвести оценку состояния и функционирования АСКУ с использованием положений теории информации.

9.2. Оценка степени неопределенности состояния объекта контроля

При контроле и управлении объект можно рассматривать как многомерную динамическую систему, на которую наряду с закономерными и случайными управляющими воздействиями или сигналами контроля влияют различные случайные помехи в виде внешних и внутренних возмущений. Состояние такой системы определяется некоторыми выходными параметрами, определенным образом

связанными с воздействиями на систему через вектор-оператор системы.

В связи со случайным характером различных воздействий и возмущений выходные параметры объекта будут также случайными функциями времени.

Полными вероятностными характеристиками как выходных параметров, так и самого объекта, являются дифференциальные многомерные законы распределения, а также уравнения для определения вероятности состояния выходов дискретных систем при различных вероятностях состояния входных сигналов. Однако они не дают интегральной качественной и количественной оценки неопределенности объекта при контроле и управлении, а также интегральной оценки изменения неопределенности объекта в процессе контроля и управления. Этот недостаток устраняется при использовании информационных характеристик объекта в процессе контроля и управления.

Для интегральной оценки неопределенности объекта с непрерывным множеством состояний в процессе контроля и управления удобно применять дифференциальную энтропию состояния объекта

$$H(X) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX, \qquad (9.1)$$

где $X(x_1, x_2, \ldots, x_m)$ — вектор выходных параметров (координат) объекта.

При независимых выходных координатах энтропия объекта равна сумме частных энтропий, обусловленных неопределенностью отдельных координат

$$H(X) = \sum_{i=1}^{m} H(x_i). \qquad (9.2)$$

Энтропия объекта с двумя дискретными состояниями (исправное и неисправное) может быть представлена также в виде

$$H(X) = -\{P_{\text{о.и}} \log_2 P_{\text{о.и}} + (1 - P_{\text{о.и}}) \log_2 (1 - P_{\text{о.и}})\}, \qquad (9.3)$$

где $P_{\text{о.и}}$ — вероятность исправного состояния объекта.

Вероятность исправного состояния объекта связана с многомерной плотностью распределения вероятности $w(X)$ соотношением

$$P_{\text{о.и}} = \int \cdots \int_{V_x} w(X)\, dX, \qquad (9.4)$$

где V_x — область допустимых значений вектора X.

Случайный разброс параметров объекта обусловлен большим количеством факторов технологического и эксплуатационного характера. Среди этих факторов обычно трудно выделить преоблада-

ющий. В такой ситуации согласно центральной предельной теореме Ляпунова закон распределения параметров очень близок к нормальному. Опираясь на это положение, закон распределения выходных параметров $w(x_i)$ в большинстве случаев можно принять нормальным

$$w(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(x_i - m_i)^2}{2\sigma_i^2}\right\}, \tag{9.5}$$

где m_i и σ_i — математическое ожидание и среднеквадратическое отклонение параметра x_i.

Тогда многомерная плотность распределения вектора X определяется соотношением [14]

$$w(X) = \frac{1}{\sigma_1\sigma_2 \ldots \sigma_m \sqrt{(2\pi)^m D}} \exp\left\{-\frac{1}{2D}\sum_{i=1}^{m}\sum_{k=1}^{m} D_{ik}\frac{x_i - m_i}{\sigma_i} \times \right.$$

$$\left. \times \frac{x_k - m_k}{\sigma_k}\right\}, \tag{9.6}$$

где $D = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{vmatrix}$ — определитель m-го порядка коэффициента корреляции;

D_{ik} — алгебраическое дополнение элемента r_{ik} в определителе D.

При статистической независимости координат x_i выражение (9.6) приводится к виду

$$w(X) = \frac{1}{\sigma_1\sigma_2 \ldots \sigma_m \sqrt{(2\pi)^m}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{m}\frac{(x_k - m_k)^2}{\sigma_k^2}\right\}. \tag{9.7}$$

Энтропия выходных параметров при нормальном законе распределения последних

$$h(x_i) = \log_2 \sigma_i \sqrt{2\pi e},$$

откуда среднеквадратическое отклонение параметра

$$\sigma_i = \frac{2^{h(x_i)}}{\sqrt{2\pi e}}. \tag{9.8}$$

При статистической независимости выходных параметров объектов

$$P_{\text{о.н}} = \prod_{i=1}^{m} p_{\text{н}i}, \tag{9.9}$$

где $p_{\text{н}i}$ — вероятность нахождения в пределах допуска i-го выходного параметра.

Тогда с учетом (9.5), (9.8) и (9.9) получим

$$p_{\text{о.и}} = 2^{\sum\limits_{i=1}^{m} h(x_i)} \sqrt{e^m} \int \cdots \int\limits_{V_x} \exp\left\{-\pi e \sum_{i=1}^{m} \frac{(x_i - m_i)^2}{2^{h(x_i)}}\right\} dx_1\, dx_2 \ldots dx_m. \tag{9.10}$$

Выражение (9.10) показывает зависимость вероятности исправного состояния объекта от энтропии выходных параметров. При наличии статистических зависимостей между выходными параметрами данное выражение оказывается достаточно сложным и здесь не приводится.

9.3. Информационная оценка точности результата контроля

Контроль — это одна из разновидностей процесса получения, преобразования, передачи и накопления информации. Под контролем во всех случаях практики понимается процесс получения человеком или машиной информации о действительном состоянии объекта контроля.

Информация, получаемая в результате контроля, количественно оценивается уменьшением энтропии от значения $H(X)$, которое характеризует неопределенность объекта контроля перед контролем, до значения $H(X/Y)$, которое остается после получения результата контроля Y, т. е. как

$$I = H(x) - H(X/Y). \tag{9.11}$$

Апостериорная энтропия $H(X/Y)$ определяется распределением вероятностей погрешностей контроля. Так как апостериорная энтропия $H(X/Y)$ обусловливает уменьшение количества информации, получаемой при контроле, то говорят, что погрешности контроля оказывают дезинформационное действие.

Дезинформационное действие погрешности в общем случае определяется законом распределения последней, причем при различных законах распределения получаются различные значения $H(X/Y)$ даже в том случае, если дисперсии погрешностей одинаковы.

Так, если погрешность распределена равномерно в пределах $+\Delta_z \div -\Delta_z$, то энтропия погрешности (см. гл. IV, пример 4.4)

$$H(X/Y) = \log_2 2\Delta_z. \tag{9.12}$$

Дисперсия погрешности с равномерным распределением

$$\sigma_z^2 = \int\limits_{-\Delta_z}^{\Delta_z} z^2 \frac{1}{2\Delta_z}\, dz = \frac{\Delta_z^2}{3}. \tag{9.13}$$

Подставив значение Δ_z из (9.13) в (9.12), получим

$$H(z) = H(X/Y) = \log_2 2\sqrt{3}\sigma_z. \tag{9.14}$$

Если погрешность распределена по нормальному закону, то энтропия, как известно (см. гл. IV, пример 4.3),

$$H(z) = H(X/Y) = \log_2 \sqrt{2\pi e}\,\sigma_z. \tag{9.15}$$

Из сравнения выражений (9.14) и (9.15) видно, что при равных σ_z дезинформационное действие нормально распределенных погрешностей больше, чем равномерно распределенных. В связи с этим указано [51] на несовершенство используемых до настоящего времени характеристик точности контроля и введено понятие так называемого энтропийного значения погрешности контроля. В качестве энтропийного значения погрешности принято значение погрешности с равномерным законом распределения, которое вносит такое же дезинформационное действие, что и погрешность с данным законом распределения вероятностей.

Энтропийное значение погрешности $\Delta_э$ с произвольным законом распределения $w(z)$ может быть получено из равенства

$$\log_2 2\Delta_э = -\int_{-\infty}^{\infty} w(z)\log_2 w(z)\,dz = H(z),$$

откуда

$$\Delta_э = \frac{1}{2}2^{H(z)} = 2^{H(z)-1}. \tag{9.16}$$

Энтропийное значение погрешности связано со среднеквадратическим значением соотношения

$$\Delta_э = k_э\sigma_z, \tag{9.17}$$

где $k_э$ — энтропийный коэффициент, величина которого определяется законом распределения погрешностей.

Наибольшей энтропией при заданном значении дисперсии обладает нормальное распределение [23], поэтому для него энтропийный коэффициент $k_э$ наибольший и равен 2,07. Любое другое распределение имеет меньшее значение коэффициента $k_э$. В частности, для равномерного распределения $k_э = 1{,}73$ [51].

Следует заметить, что изложенное относится к случаю, когда контроль параметров осуществляется по критерию количественной оценки, т. е. по существу сводится только к измерению величины параметров. В тех же случаях, когда контроль осуществляется по допусковому критерию, т. е. когда целью контроля является установление, находится или не находится величина контролируемого параметра в пределах допуска, энтропия погрешности контроля должна определяться соотношением

$$h(X/Y) = -\{p_{ош}\log_2 p_{ош} + (1-p_{ош})\log_2(1-p_{ош})\}, \tag{9.18}$$

где $p_{ош}$ — общая безусловная вероятность ошибочного решения относительно состояния контролируемого параметра, определяемая выражением (8.28).

Ошибки первого и второго рода, входящие в выражение (8.28), можно определить соответственно формулами [53]:

$$\alpha = \int_{-\infty}^{x_1} w(x)\left[\int_{x_1-x}^{x_2-x} w(\xi)\,d\xi\right]dx + \int_{x_2}^{\infty} w(x)\left[\int_{x_1-x}^{x_2-x} w(\xi)\,d\xi\right]dx;$$

$$\beta = \int_{x_1}^{x_2} w(x)\left[\int_{-\infty}^{x_1-x} w(\xi)\,d\xi + \int_{x_2-x}^{\infty} w(\xi)\,d\xi\right]dx,$$

где x_1 и x_2 — допуски на контролируемый параметр; $w(x)$ и $w(\xi)$ — законы распределения вероятностей контролируемого параметра x и погрешностей ξ контроля.

Очевидно и в этом случае может быть введено энтропийное значение погрешности контроля, которое по дезинформационному действию эквивалентно погрешности $p_{ош}$. Причем величина этой погрешности может быть определена по формуле (9.16), если в показатель степени числа 2 подставить энтропию, определяемую выражением (9.18).

9.4. Информационная способность устройства контроля и информационный к. п. д. процесса контроля

Под информационной способностью устройства контроля можно понимать эквивалентное число различимых интервалов, определяющее получаемое от устройства контроля количество информации [51].

Таким образом, согласно определению информационная способность устройства контроля N может быть определена из соотношения

$$I = \log_2 N, \tag{9.19}$$

где I — количество информации, получаемой при контроле.

Из (9.19) получим следующее выражение для информационной способности:

$$N = 2^I. \tag{9.20}$$

Если во всем диапазоне измерений энтропийная погрешность постоянна, то количество информации I определяется формулой (9.11). Если же $\Delta_э$ переменна, то в формулу (9.11) в качестве второго члена должно быть подставлено средневзвешенное, в соответствии с распределением плотности $w(x)$, значение условной энтропии

$$H_{ср}(x/y) = -\int_{-\infty}^{\infty} w(x)\,H(x/y)\,dx. \tag{9.21}$$

В реальных условиях контролируемый параметр x подвержен различным внешним случайным возмущениям. Имеющиеся в результате этого флюктуации параметра x характеризуют обычно среднеквадратическим значением σ_ξ.

По аналогии с энтропийной погрешностью устройства контроля можно пользоваться понятием энтропийной погрешности параметра x

$$\Delta_\xi = k_\xi \sigma_\xi, \tag{9.22}$$

где k_ξ — энтропийный коэффициент, величина которого определяется законом распределения флюктуаций $w(\xi)$.

Следовательно, можно утверждать, что контролируемый параметр x_1, поступающий на вход устройства контроля, несет количество информации

$$I_x = H(x) - \int_{-\infty}^{\infty} w(\xi) \log_2 w(\xi)\, d\xi \tag{9.23}$$

и характеризуется информационной способностью

$$N_x = 2^{I_x}. \tag{9.24}$$

Вследствие того, что устройство контроля обладает погрешностями, количество информации I_1, получаемое на выходе устройства контроля, меньше информации I_x, которую несет параметр x. Вследствие этого процесс контроля сопровождается потерей информации

$$\Delta I = I_x - I_1. \tag{9.25}$$

Величина ΔI является информационной оценкой совершенства процесса контроля. Обычно пользуются относительной величиной [51]

$$\eta_и = \frac{I_1}{I_x} = \frac{\log_2 N_1}{\log_2 N_x}, \tag{9.26}$$

получившей наименование информационного к. п. д. процесса контроля.

В формулах (9.25) и (9.26) количество информации I_1, реально получаемое в результате контроля, определяется суммарной погрешностью контролируемого параметра и устройства контроля. Однако в подавляющем большинстве практических случаев погрешность, вносимая устройством контроля, существенно превышает флюктуации контролируемого параметра. Поэтому в суммарной погрешности слагаемым, обусловленным флюктуациями контролируемой величины, можно пренебречь с сохранением весьма высокой степени точности.

9.5. Пропускная способность процесса контроля

Пропускная способность процесса контроля $V_к$ является динамической характеристикой и выражает максимально возможное количество информации, которое можно получить в процессе контроля в единицу времени [52],

$$V_к = \frac{I_{макс}}{\overline{T}_к}, \qquad (9.27)$$

где $I_{макс}$ — максимальное количество информации, которое можно получить в среднем при контроле одного объекта; $\overline{T}_к$ — математическое ожидание времени, затрачиваемого на контроль одного объекта.

Если объект контроля имеет один параметр или несколько параметров, контролируемых параллельно, то $\overline{T}_к$ определяется временем контроля одного параметра.

При последовательном контроле нескольких параметров объекта контроль обычно прекращается после обнаружения первого отказа. В связи с этим при последовательном контроле время контроля будет определяться вероятностью обнаружения отказов отдельных параметров объекта

$$\overline{T}_к = \sum_{i=1}^{m} p_{i-1} t_{кi}, \qquad (9.28)$$

где $t_{кi}$ — время контроля i-го параметра объекта; p_{i-1} — априорная вероятность того, что при контроле $(i-1)$-го параметра отказ объекта не будет обнаружен.

При заданном законе распределения погрешностей контроля количество информации, получаемое в процессе контроля, будет максимально в случае статистической независимости и при равномерном законе распределения контролируемых параметров. В этом случае количество информации, получаемое при контроле,

$$I_{макс} = \sum_{i=1}^{m} [H_{макс}(x_i) - H(\xi_i)] = \sum_{i=1}^{m} \log_2 \frac{(x_{i_2} - x_{i_1})}{2\Delta_{эi}}, \qquad (9.29)$$

где x_{i_1} и x_{i_2} — пределы изменения i-го контролируемого параметра.

Подставив из (9.28) и (9.29) в (9.27) значения $\overline{T}_к$ и $I_{макс}$, получим

$$V_к = \frac{\sum\limits_{i=1}^{m} \log_2 \frac{x_{i_2} - x_{i_1}}{2\Delta_{эi}}}{\sum\limits_{i=1}^{m} p_{i-1} t_{кi}}. \qquad (9.30)$$

Если при контроле решается двухальтернативная задача, то максимальное количество информации будет получено при одинаковой вероятности «норма» и «не норма» контролируемых параметров. При этом априорная вероятность контролируемых параметров равна 1 дв. ед. и выражение (9.30) принимает вид

$$V_{\text{к}} = \frac{m \log_2 \frac{1}{\Delta_{\text{э}}}}{\sum_{i=1}^{m} p_{i-1} t_{\text{к}i}}. \tag{9.31}$$

Из выражений (9.30) и (9.31) видно, что параметр $V_{\text{к}}$ является характеристикой быстродействия и точности процесса контроля. При заданной пропускной способности контроля точность контроля может быть повышена в обмен на снижение быстродействия, и наоборот.

9.6. Энтропия и информация в системах автоматического управления

Понятия энтропии и информации являются удобными обобщенными характеристиками при описании сложных систем автоматического управления.

Любой процесс, подлежащий автоматическому управлению, можно охарактеризовать совокупностью координат $x_1, x_2, \ldots, x_m$. Эти координаты практически всегда имеют разброс относительно номинальных значений. Этот разброс обусловлен многочисленными факторами и имеет случайный характер. Для оценки неопределенности процесса может быть использована энтропия распределения вероятностей координат управляемого процесса [50], выражаемая такой же формулой, что и энтропия, рассмотренная ранее в системах передачи информации,

$$H(X) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w(X) \log_2 w(X)\, dX. \tag{9.32}$$

Несмотря на общность математических выражений, энтропия процесса отличается от информационной энтропии по существу. В системах передачи информации после получения информации энтропия объекта уменьшается. Энтропия же процесса после получения информации о его состоянии наблюдателем не меняется. Следовательно, энтропию распределения вероятностей координат процесса нельзя изменить измерением этих координат, для изменения энтропии процесса необходимо управление или восстановление.

Если проводить аналогию с процессом контроля параметров, как процессом передачи информации, то введенное выше понятие энтропии процесса соответствует при контроле энтропии парамет-

ров за счет их флюктуации вследствие внешних возмущающих воздействий.

Энтропия процесса может быть изменена только за счет воздействия на этот процесс. Основным назначением автоматизированного управления процессом является уменьшение отклонений процесса от заданного, уменьшение неопределенности протекания процесса, т. е. уменьшение энтропии процесса.

Для иллюстрации вышеизложенного рассмотрим следующий пример [50].

Генератор переменного тока имеет при некоторых условиях разброс напряжения $\sigma_{0v} = 50$ В и частоты $\sigma_{0f} = 50$ Гц. При включении системы регулирования напряжения и частоты разброс уменьшается до $\sigma_{1v} = 2$ В и $\sigma_{1f} = 1$ Гц. Найти уменьшение энтро-

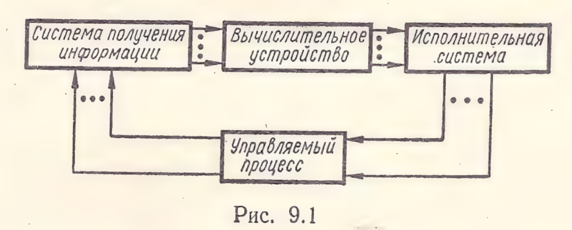


Рис. 9.1

пии координат процесса при переходе от неуправляемого процесса к управляемому при условии независимости и нормального распределения вероятностей координат v и f.

Общее уменьшение энтропии будет равно сумме уменьшений энтропий координат

$$H_0 - H_1 = H_0(v) - H_1(v) + H_0(f) - H_1(f) =$$

$$= \log_2 \frac{\sigma_{ov}}{\sigma_{1v}} + \log_2 \frac{\sigma_{of}}{\sigma_{1f}} = -\log_2 25 + -\log_2 10 = -7{,}97 \text{ дв. ед.}$$

Для процессов управления характерны замкнутые контуры циркуляции информации. Схема такого контура изображена на рис. 9.1. Параметры процесса $x_1, x_2, \ldots, x_m$ измеряются системой получения информации. Полученная информация обрабатывается вычислительным устройством. Обработанная информация поступает в исполнительную систему, формулирующую управляющие воздействия z_i, которые воздействуют на управляемый процесс в направлении уменьшения его энтропии.

Управление процессом может быть непрерывным или дискретным во времени. Ограничимся рассмотрением дискретного управления, характерного для контуров с цифровыми управляющими машинами [9, 32, 50].

Управляющие воздействия поступают на управляемый процесс через интервалы времени τ.

Пусть до передачи управляющих воздействий начальная энтропия процесса равна $H_0(X)$. После включения системы управления

энтропия процесса уменьшается до значения $H_\xi(X)$, определяемого воздействием различного рода возмущений в контуре управления (погрешностей измерителей, преобразователей и исполнительных устройств, внешних возмущений и пр.). Количество информации, внесенное при этом в контур управления,

$$I(X) = H_0(X) - H_\xi(X). \tag{9.33}$$

Этот процесс будет происходить за определенное количество тактов работы устройства управления. После подачи каждого управляющего воздействия энтропия процесса уменьшается на определенную величину. В интервалах между воздействиями энтропия процесса растет.

Количество информации $I_k(X)$, вносимое в контур управления в k-м такте работы,

$$I_k(X) = H_{k-1}(X) - H_k(X) + \Delta H_k(X), \tag{9.34}$$

где $H_{k-1}(X)$ — энтропия процесса в конце $(k-1)$ такта работы; $H_k(X)$ — энтропия процесса в конце k-го такта работы; $\Delta H_k(X)$ — прирост энтропии процесса за интервал дискретности τ вследствие действия возмущений.

Если за время переходного периода, в течение которого энтропия процесса уменьшается от значения $H_0(X)$ до значения $H_\xi(X)$, совершено n тактов работы устройства управления, то общее количество информации, внесенное в контур управления за это время,

$$I_{\text{пер}}(X) = \sum_{k=1}^{n} [H_{k-1}(X) - H_k(X)] + \sum_{k=1}^{n} \Delta H_k(X). \tag{9.35}$$

В установившемся режиме будет соблюдаться условие

$$\begin{gathered} H_{k-1}(X) = H_k(X); \\ I_k(X) = \Delta H_k(X), \end{gathered} \tag{9.36}$$

т. е. количество информации, передаваемое через контур управления за интервал дискретности, будет равно приросту энтропии процесса за счет действия возмущений.

Суммарный прирост количества информации за время переходного периода равен суммарному уменьшению энтропии процесса

$$I(x) = \sum_{k=1}^{n} [H_{k-1}(X) - H_k(X)]. \tag{9.37}$$

Характер изменения энтропии процесса X в переходном и установившемся режимах изображен на рис. 9.2.

Рассмотрим следующий пример [50].

Предположим, что необходимо построить систему стабилизации летательного аппарата, в контур которой включена ЦВМ, обеспечивающая в стационарном режиме:

точность стабилизации угла тангажа $\sigma_\psi = 1°$;
точность стабилизации угла рыскания $\sigma_\beta = 0,5°$;
точность стабилизации угла вращения $\sigma_\alpha = 0,3°$;
точность стабилизации дальности полета $\sigma_н = 50$ м;
точность стабилизации ЦМ по нормали $\sigma_x = 30$ м.

Требуется, чтобы система стабилизации после ее включения выводила объект в указанный стационарный режим не более чем за 30 с при следующем начальном разбросе координат (до включения системы):

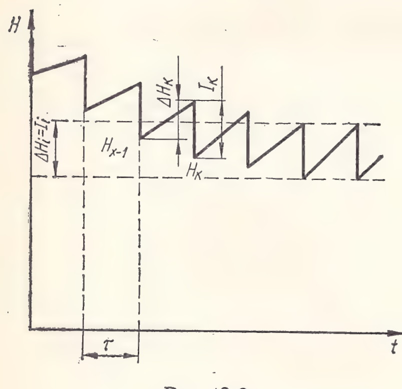


Рис. 9.2

$$\sigma_{0\psi} = 8°;\ \sigma_{0\beta} = 10°;\ \sigma_{0\alpha} = 20°;$$
$$\sigma_{0н} = 3\ \text{км};\ \sigma_{0x} = 2\ \text{км}.$$

В случае внезапного прекращения управления в стационарном режиме на 1 с, отклонения нарастают до следующих значений:

$$\sigma'_\psi = 1,5°;\ \sigma'_\beta = 5°;\ \sigma'_\alpha = 10°;$$
$$\sigma'_н = 55\ \text{м};\ \sigma'_x = 33\ \text{м}.$$

Для простоты закон распределения вероятностей выходных координат принимается нормальным, координаты — независимыми, интервал дискретности управления — одинаковым для всех координат и равным $\tau = 1$ с, а также то, что сами координаты хранятся в специальных запоминающих устройствах и в примере не рассматриваются.

1) Определить количество информации, которое необходимо передавать через полный многомерный контур управления в стационарном режиме в течение интервала дискретности.

2) Оценить общее количество информации, которое необходимо определить через контур управления в переходном режиме.

3) Оценить нижнюю границу емкости памяти ЦВМ.

Так как количество информации, передаваемое через контур управления за интервал дискретности в установившемся режиме, равно приросту энтропии процесса за счет действия возмущений, то ответ на первый вопрос будет следующим:

$$I_k = \log_2 \frac{\sigma'_\psi}{\sigma_\psi} + \log_2 \frac{\sigma'_\beta}{\sigma_\beta} + \log_2 \frac{\sigma'_\alpha}{\sigma_\alpha} + \log_2 \frac{\sigma'_н}{\sigma_н} + \log_2 \frac{\sigma'_x}{\sigma_x} =$$
$$= \log_2 1,5 + \log_2 10 + \log_2 33,3 + \log_2 1,1 + \log_2 1,1 = 9,239\ \text{дв. ед.}$$

Общее количество информации, передаваемое через контур управления в переходном режиме, равно сумме общего изменения

энтропии процесса за счет введения управления и прироста энтропии за счет действия возмущений. Так как за время переходного периода совершается $m = 30$ тактов управления, то

$$I_{\text{пер}} = H_0 - H_1 + 30 I_k = \log_2 \frac{\sigma_{0\psi}}{\sigma_\psi} + \log_2 \frac{\sigma_{0\beta}}{\sigma_\beta} + \log_2 \frac{\sigma_{0\alpha}}{\sigma_\alpha} +$$
$$+ \log_2 \frac{\sigma_{0н}}{\sigma_н} + \log_2 \frac{\sigma_{0x}}{\sigma_x} = \log_2 8 + \log_2 20 + \log_2 66{,}6 +$$
$$+ \log_2 15 + \log_2 20 + 30 I_k = 298{,}87 \text{ дв. ед.}$$

Нижняя граница емкости памяти ЦВМ

$$Q > \frac{I_{\text{пер}}}{m} = \frac{298{,}87}{30} = 9{,}9 \text{ дв. ед./такт.}$$

9.7. Оценка эффективности АСКУ

С информационной точки зрения эффективность АСКУ целесообразно оценивать количеством информации, получаемой системой в определенный интервал времени с учетом экономических затрат. В связи с этим эффективность АСКУ следует оценивать величиной

$$K = \frac{I(X)}{C}, \qquad (9.38)$$

где $I(X)$ — количество информации, получаемое АСКУ в определенный интервал времени; C — стоимость АСКУ.

Неудобство критерия эффективности, выражаемого формулой (9.38), заключается в его ненормированности. Критерием, лишенным данного недостатка, является обобщенный статистический критерий [49, 54, 55, 56], выражающий отношение эффективностей реальной и потенциальной АСКУ

$$Э = \frac{K_р}{K_п}, \qquad (9.38)$$

где $K_р$ — эффективность реальной АСКУ; $K_п$ — эффективность потенциальной (идеальной) АСКУ.

Потенциальная АСКУ обеспечивает получение максимально возможного количества информации. Это будет иметь место при максимуме априорной и минимуме апостериорной энтропии объекта контроля. Следовательно, для потенциальной системы должно быть равенство априорных вероятностей контролируемых параметров

$$p(x_i) = q(x_i),$$

где $p(x_i)$ — вероятность состояния «норма» i-го параметра*; $q(x_i)$ — вероятность состояния «не норма» i-го параметра.

---

* Имеется в виду, что практически $p(x_i) \geqslant 0{,}5$.

При этом априорная энтропия объекта будет равна

$$H_0(x) = -\sum_{i=1}^{m} [p(x_i) \log_2 p(x_i) + q(x_i) \log_2 q(x_i)] = m \text{ дв. ед.} \quad (9.39)$$

Решающее устройство потенциальной системы работает по критерию идеального наблюдателя, минимизирующего суммарную ошибку принятия решения

$$q(x_i)\,\alpha_i + p(x_i)\,\beta_i = \text{мин}, \quad (9.40)$$

где α_i и β_i — ошибки первого и второго рода.

В результате осуществления контроля потенциальная система переводит состояние объекта в некоторое более определенное состояние, характеризуемое в соответствии с формулой Байеса апостериорной вероятностью исправного состояния

$$p(x_i/y_i) = \frac{p(x_i)[1-\alpha_i]}{p(x_i)[1-\alpha_i] + q(x_i)\beta_i} \quad (9.41)$$

и энтропией

$$H_1(x_i) = -p(x_i/y_i) \log_2 p(x_i/y_i) - [1 - p(x_i/y_i)] \log_2 [1 - p(x_i/y_i)]. \quad (9.42)$$

Общее количество информации, получаемое потенциальной системой

$$I_{\text{п}}(x) = I_{\text{макс}} = \sum_{i=1}^{m} H_0(x_i) - \sum_{i=1}^{m} H_1(x_i). \quad (9.43)$$

Потенциальная система является идеальной не только в смысле максимума количества получаемой информации, она является идеальной также в смысле простоты, так как в ней не предусмотрено резервирования, доработок для получения нужного быстродействия, объема, веса и т. п.

Следовательно, потенциальная система при выполнении указанных условий будет иметь минимальную стоимость $C_{\text{мин}}$.

Эффективность потенциальной системы оценивается коэффициентом

$$K_{\text{п}} = \frac{I_{\text{макс}}}{C_{\text{мин}}}. \quad (9.44)$$

Математическая модель реальной АСКУ строится аналогично математической модели потенциальной системы, однако при этом принимаются реальные законы распределения вероятностей состояний контролируемых параметров и различные алгоритмы работы решающих устройств, обусловленные выбранным критерием оценки параметров.

Общая стоимость C информации $I_р$ с учетом затрат на получение нужной надежности, быстродействия, объема, веса и т. п. в реальной АСКУ

$$C = \sum_{i=1}^{m} C_i > C_{мин}. \qquad (9.45)$$

Эффективность работы реальной АСКУ

$$K_р = \frac{I_р}{C} \leqslant K_п. \qquad (9.46)$$

Достоинством обобщенного статистического критерия (9.38), полученного на основе потенциальной и реальной математических моделей АСКУ, является полная наглядность, сравнительная простота и общность, позволяющая одним числом характеризовать как всю систему, так и по частям, включающим сложные и простые устройства. При этом диапазон изменения критерия для практических систем

$$0 \leqslant Э \leqslant 1. \qquad (9.47)$$

Несовершенная АСКУ имеет $Э$, близкий к нулю, совершенная — к единице. Дезинформирующая АСКУ имеет $Э$ меньше нуля.

9.8. Информационное обоснование оптимального алгоритма контроля и поиска неисправностей

Всякий сложный объект можно представить в виде системы, состоящей из нескольких динамических последовательно соединенных звеньев, которые, в свою очередь, состоят из элементов.

При эксплуатации таких систем различают два специфических алгоритма [57]:

алгоритм контроля, предназначенного для установления факта исправности или неисправности системы;

алгоритм поиска неисправностей в случае установления факта неисправности системы.

Определение оптимальных алгоритмов контроля и поиска неисправностей в системе можно произвести на основе теории информации, теории игр, теории массового обслуживания и теории последовательного анализа Вальда.

При определении указанных алгоритмов на основе теории информации можно использовать два различных критерия:

критерий максимального количества информации в единицу времени, с применением которого контроль и поиск неисправности начинается с опыта, дающего максимальное количество информации в единицу времени;

критерий максимальной ценности информации, с применением которого контроль и поиск неисправности начинается с опыта, дающего максимальную ценность информации.

Критерий максимального количества информации целесообразно использовать при автоматическом контроле и поиске неисправностей.

Критерий максимальной ценности информации следует использовать при разработке инструкций по эксплуатации оборудования обслуживающим персоналом.

Рассмотрим методику выбора оптимальных алгоритмов контроля и поиска неисправностей в системе, основанной на применении критерия максимального количества информации. При этом предполагается, что система контроля работает без ошибок.

Неопределенность состояния контролируемой системы до начала контроля и поиска неисправностей характеризуется ее полной энтропией $H(X)$.

Процесс контроля представляет собой сложный опыт A_k, состоящий не более чем из k испытаний a.

Среднее количество информации, получаемое при каждом испытании a_i относительно состояния системы X,

$$I(a_i,\ X) = H(X) - H_{a_i}(X),$$

где $H_{a_i}(X) = p(A_{i_1}) H_{A_{i_1}}(X) + \cdots + P(A_{ij}) H_{A_{ij}}(X)$ — средняя условная энтропия состояния системы после осуществления опыта a_i; $p(A_{ij})$ — вероятность отдельных исходов A_{ij} опыта a_i.

Когда исход сложного опыта A_k однозначно определяет состояние системы, характеризуемое вектором X, тогда энтропия системы после осуществления этого опыта $H(X) = 0$, а среднее количество информации о состоянии системы, полученное при осуществлении опыта A_k,

$$I(A_k,\ X) = H(X).$$

Когда исход сложного опыта не полностью раскрывает работоспособность состояния системы, тогда

$$H_{A_k}(X) \neq 0 \text{ и } I(A_k,\ X) < H(X).$$

Основа оптимального алгоритма контроля состоит в выборе такого минимального количества опытов a, являющихся составляющими сложного опыта A_k, каждый из которых давал бы максимальное количество информации.

Процесс контроля должен вестись по алгоритму, составленному на основе критерия

$$I(a_i,\ X) = H(X) - H_{ai}(X) = I_{\text{макс}},$$

т. е. процесс контроля должен начинаться с опыта, дающего максимальное количество информации.

После осуществления опыта a_1 система будет обладать энтропией некоторого нового состояния

$$H_{a1}(X).$$

Предположим, что состояние системы до начала процесса контроля характеризуется состояниями: система исправна с вероятностью P; система неисправна с вероятностью $1 - P$.

Кроме того, для упрощения задачи, а также с учетом опыта эксплуатации предположим, что система не работает вследствие отказа только одного элемента из m, «соединенных» цепочкой, имеющего условную вероятность отказа q_k, определяемую на основе статистических сведений о надежности элементов.

Таким образом, отказы элементов системы предполагаются несовместимыми, а для полной группы элементов справедливо равенство (при наличии отказа)

$$\sum_{k=1}^{m} q_k = 1.$$

Энтропия системы

$$H(X) = -\left\{P \log_2 P + (1 - P) \log_2 (1 - P) + (1 - P) \sum_{k=1}^{m} q_k \log_2 q_k\right\}. \tag{9.48}$$

Предположим, что при осуществлении опыта a_1 контролируется n элементов системы, а также, что имеется два исхода опыта:

контролируемая часть системы исправна (вероятность этого исхода $\left[1 - (1 - P) \sum_{k=1}^{n} q_k\right]$);

контролируемая часть системы неисправна (вероятность этого исхода $(1 - P) \sum_{k=1}^{n} q_k$).

Средняя условная энтропия состояния системы после осуществления первого опыта

$$H_{a_1}(X) = -\{P \log_2 P + (1 - P) \log_2 (1 - P) + (1 - P) \sum_{i=1}^{n} q_k \log_2 q_k - \\ - [1 - \sum_{k=1}^{n} q_k (1 - P)] \log_2 [1 - \sum_{k=1}^{n} q_k (1 - P)] - \\ - \sum_{k=1}^{n} q_k (1 - P) \log_2 (1 - P) \sum_{k=1}^{n} q_k\}. \tag{9.49}$$

Количество информации о состоянии системы, получаемое при осуществлении опыта

$$I(a_1, X) = -\{[1 - \sum_{k=1}^{n} q_k (1 - P)] \log_2 [1 - \sum_{k=1}^{n} q_k (1 - P)] + \\ + [\sum_{k=1}^{n} q_k (1 - P)] \log_2 [\sum_{k=1}^{n} q_k (1 - P)]\}. \tag{9.50}$$

Задача состоит в том, чтобы при первом этапе контроля выбрать такое количество контролируемых элементов, при котором количество получаемой информации было бы максимальным.

Из формулы (9.50) видно, что $I(a_1, X)$ зависит от P и $\sum_{k=1}^{n} q_k$ и является максимальным при выполнении условия

$$1-\sum_{k=1}^{n} q_k (1-P) = \frac{1}{2},$$

или

$$\sum_{k=1}^{n} q_k = \frac{1}{2(1-P)}. \qquad (9.51)$$

Рассмотрим, как изменяется n в зависимости от вероятности исправного состояния объекта P при соблюдении условия (9.51).

При $P=0$ должно быть справедливо равенство $\sum_{k=1}^{n} q_k = 0{,}5$. При равенстве вероятностей отказов элементов $q_1 = q_2 = \cdots q_m = \frac{1}{m}$ получим

$$n = \frac{m}{2}, \qquad (9.52)$$

т. е. первым этапом контроля должна быть охвачена половина элементов объекта.

С ростом P растет n. При $P = 0{,}5$ $n = m$, т. е. контролем должны быть на первом этапе охвачены все элементы. С дальнейшим ростом P увеличивается $\sum_{k=1}^{n} q_k$. Но так как в реальных условиях сумма $\sum_{k=1}^{n} q_k$ не может быть больше единицы, то при $P > 0{,}5$ принимается $\sum_{k=1}^{n} q_k = 1$ (рис. 9.3) или $n = m$.

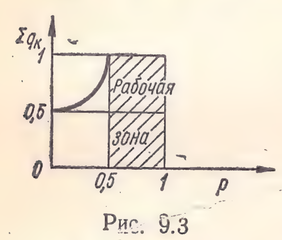


Рис. 9.3

Так как для реальных систем вероятность исправного состояния системы $P > 0{,}5$, то при выборе первого опыта необходимо алгоритм построить так, чтобы охватить контролем $m = n$ элементов системы.

Если $P < 0{,}5$ (что практически маловероятно), то первым этапом контроля охватываются не все элементы объекта. Тогда следующий опыт выбирается на основе того же критерия, но с учетом того, что состояние системы характеризуется энтропией (9.49). Опыты ведутся до тех пор, пока энтропия системы не станет равной нулю.

Алгоритм поиска неисправности в системе составляется при следующих предположениях:

известно, что система неисправна;

в системе отказал только один элемент из m, «соединенных» цепочкой;

известны априорные вероятности отказа всех элементов системы.

Методика определения оптимального алгоритма поиска неисправностей такая же, как и при контроле, т. е. заключается в выборе такого количества опытов, чтобы каждый из них давал максимальное количество информации.

Пусть на первом этапе поиска неисправностей a_1 рассматривается n элементов объекта из общего числа m. При этом возможны два исхода опыта:

в рассматриваемой части объекта обнаружен отказ (вероятность этого исхода $\sum_{i=1}^{n} q_k$);

в рассматриваемой части объекта не обнаружен отказ (вероятность этого исхода $1-\sum_{k=1}^{n} q_k$).

Количество информации, получаемое при осуществлении опыта a_1

$$I(a_1,\ x) = -\left\{\sum_{k=1}^{n} q_k \log_2 \sum_{k=1}^{n} q_k + \left(1-\sum_{k=1}^{n} q_k\right) \log_2 \left(1-\sum_{k=1}^{n} q_k\right)\right\}. \tag{9.53}$$

Из формулы (9.53) видно, что для получения максимального количества информации $I(a_1,\ x)$ при осуществлении первого опыта поиска неисправности необходимо, чтобы сумма $\sum_{k=1}^{n} q_k$ была равной 1/2. Следующий опыт выбирается на основании того же критерия, но с учетом энтропии состояния системы после осуществления первого опыта a_1.

Так, например, при поиске отказа в системе, состоящей из m элементов, имеющих одинаковую вероятность отказа, необходимо в первый опыт проверить $n=\frac{m}{2}$ элементов, во второй — $\frac{n}{2}$ и т. д.

Таким образом, алгоритм поиска неисправности отличается от алгоритма контроля вследствие того, что критерии выбора первого опыта различны. Кроме того, после проведения опыта a_1 опыт a_2 в случае поиска неисправности производится исходя из опыта a_1, соответствующего неработоспособности системы, а в случае контроля — исходя из опыта a_1, соответствующего работоспособности части системы.

9.9. Оценка методов и устройств контроля динамической системы стабилизации летательного аппарата

Построение функционально-статистических математических моделей динамических систем выполнено по упрощенной модели с учетом:

линеаризации нелинейностей, входящих в систему, являющейся практически только регулятором;

нормального характера законов распределения выходных параметров и параметров элементов системы, а также равенства нулю математических ожиданий отклонений от номинальных параметров:

независимости статистических характеристик выходных параметров и параметров элементов системы от времени.

Необходимость построения более детальных математических моделей с учетом всевозможных факторов определяется характером и возможностью решаемой практической задачи.

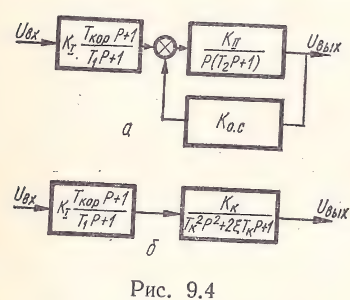


Рис. 9.4

Под выходными параметрами системы в данном примере понимаются частотные характеристики и переходная функция, а под параметрами элементов — параметры динамических звеньев системы, структурная схема которой изображена на рис. 9.4, *а*, где $k_{\rm I}$ — коэффициент усиления корректирующего звена; $T_{\text{кор}}$, T_1 — постоянные времени корректирующего звена; $k_{\rm II}$ — коэффициент усиления привода; T_2 — постоянная времени привода; $k_{\text{о. с}}$ — коэффициент усиления цепи обратной связи.

При преобразовании замкнутого контура к эквивалентному колебательному звену с параметрами

$$k_k = \frac{1}{k_{\text{о. с}}};$$

$$T_k = \sqrt{\frac{T_2}{k_{\rm II}k_{\text{о. с}}}}; \qquad (9.54)$$

$$\xi = \frac{1}{2}\cdot\frac{1}{\sqrt{T_2k_{\rm II}k_{\text{о. с}}}}$$

структурная схема имеет вид, изображенный на рис. 9.4, *б*.

Этим структурным схемам соответствует передаточная функция, являющаяся функциональной математической моделью системы,

$$W(p) = \frac{k_{\rm I}(T_{\text{кор}}p+1)}{T_1p+1}\cdot\frac{k_{\rm II}}{p(T_2p+1)+k_{\rm II}k_{\text{о. с}}} = \frac{k_{\rm I}(T_{\text{кор}}p+1)}{T_1p+1}\cdot\frac{k_k}{T_k^2p^2+2\xi T_kp+1}. \qquad (9.55)$$

Для упрощения методики определения частных производных по параметрам воспользуемся функциями чувствительности относительно параметров $k_{\rm I}$, $T_{\text{кор}}$, T_1, T_2, $k_{\rm II}$, $k_{\text{о. с}}$.

Получим

$$S_{k_{\rm I}}(p) = \frac{\partial \ln W(p)}{\partial \ln k_{\rm I}} = 1;$$

$$S_{T_{\text{кор}}}(p) = \frac{\partial \ln W(p)}{\partial \ln T_{\text{кор}}} = \frac{T_{\text{кор}}p}{T_{\text{кор}}p+1};$$

$$S_{T_1}(p)=\frac{\partial \ln W(p)}{\partial \ln T_1}=\frac{T_1p}{T_1p+1};$$

$$S_{k_{II}}(p)=\frac{\partial \ln W(p)}{\partial \ln k_{II}}=\frac{1}{1+\dfrac{k_{II}k_{o.c}}{p(T_2p+1)}};$$

$$S_{T_2}(p)=\frac{\partial \ln W(p)}{\partial \ln T_2}=-\frac{1}{1+\dfrac{k_{II}k_{o.c}}{p(T_2p+1)}}\cdot\frac{T_2p}{T_2p+1};$$

$$S_{k_{o.c}}(p)=\frac{\partial \ln W(p)}{\partial \ln k_{o.c}}=\frac{k_{II}\cdot k_{o.c}}{p(T_2p+1)}. \qquad (9.56)$$

Учитывая соотношение (9.54), функции чувствительности по параметрам замкнутого контура для последующих расчетов удобнее выразить через функции чувствительности по параметрам эквивалентного замкнутого контура колебательного звена

$$S_{k_{II}}(p)=\frac{\partial \ln W(p)}{\partial \ln k_k}\cdot\frac{\partial \ln k_k}{\partial \ln k_{II}}+\frac{\partial \ln W(p)}{\partial \ln T_k}\cdot\frac{\partial \ln T_k}{\partial \ln k_{II}}+\frac{\partial \ln W(p)}{\partial \ln \xi}\cdot\frac{\partial \ln \xi}{\partial \ln k_{II}};$$

$$S_{T_2}(p)=\frac{\partial \ln W(p)}{\partial \ln k_k}\cdot\frac{\partial \ln k_k}{\partial \ln T_2}+\frac{\partial \ln W(p)}{\partial \ln T_k}\cdot\frac{\partial \ln T_k}{\partial \ln T_2}+\frac{\partial \ln W(p)}{\partial \ln \xi}\cdot\frac{\partial \ln \xi}{\partial \ln T_2};$$

$$S_{k_{o.c}}(p)=\frac{\partial \ln W(p)}{\partial \ln k_k}\cdot\frac{\partial \ln k_k}{\partial \ln k_{o.c}}+\frac{\partial \ln W(p)}{\partial \ln T_k}\frac{\partial \ln T_2}{\partial \ln k_{o.c}}+\frac{\partial \ln W(p)}{\partial \ln \xi}\cdot\frac{\partial \ln \xi}{\partial \ln k_{o.c}}.$$

После выполнения операции дифференцирования с учетом (9.54) получим

$$S_{k_{II}}(p)=-\frac{1}{2}S_{T_k}(p)-\frac{1}{2}S_{\xi}(p);$$

$$S_{T_2}(p)=\frac{1}{2}S_{T_k}(p)-\frac{1}{2}S_{\xi}(p);$$

$$S_{k_{o.c}}(p)=-S_{k_k}(p)-\frac{1}{2}S_{T_k}(p)-\frac{1}{2}S_{\xi}(p), \qquad (9.57)$$

где

$$S_{k_k}(p)=\frac{\partial \ln W(p)}{\partial \ln k_k}=1;$$

$$S_{T_k}(p)=\frac{\partial \ln W(p)}{\partial \ln T_k}=-\frac{2T_k^2p^2+2\xi T_kp}{T_k^2p^2+2\xi T_kp+1};$$

$$S_{\xi}(p)=\frac{\partial \ln W(p)}{\partial \ln \xi}=-\frac{2\xi T_kp}{T_k^2p^2+2\xi T_kp+1}.$$

Полученные функции чувствительности по шести параметрам системы, выраженные через функции чувствительности по параметрам элементарных звеньев, позволяют использовать при расчете веса отклонения характеристики элементарных звеньев, определяемые заранее.

Информационная оценка метода и устройства контроля динамической системы, основанного на измерении амплитудно-частотной характеристики (АЧХ)

Определим веса относительных отклонений АЧХ для параметров системы

$$w_{\Delta Ak_{\mathrm{I}}}(\omega) = Re\,[S_{k_{\mathrm{I}}}(j\omega)];$$
$$w_{\Delta AT_{\text{кор}}}(\omega) = Re\,[S_{T_{\text{кор}}}(j\omega)];$$
$$w_{\Delta AT_1}(\omega) = Re\,[S_{T_1}(j\omega)]; \qquad (9.58)$$
$$w_{\Delta Ak_{\mathrm{II}}}(\omega) = -\frac{1}{2} Re\,[S_{T_k}(j\omega) + S_{\xi}(j\omega)];$$
$$w_{\Delta AT_2}(\omega) = \frac{1}{2} Re\,[S_{T_k}(j\omega) - S_{\xi}(j\omega)];$$
$$w_{\Delta Ak_{\text{о. с}}}(\omega) = Re\left[S_{k_k}(j\omega) - \frac{1}{2} S_{T_k}(j\omega) - \frac{1}{2} S_{\xi}(j\omega)\right].$$

В результате выполнения расчетов по системе уравнений (9.58) получим значения весов относительных отклонений АЧХ, вызванных отклонениями параметров системы (рис. 9.5).

Энтропия АЧХ, обусловленная отдельными параметрами системы, позволяет определить параметры системы, вносящие наибольшую неопределенность в частотном диапазоне. Эту характеристику возможно определить по энтропии элементарных звеньев.

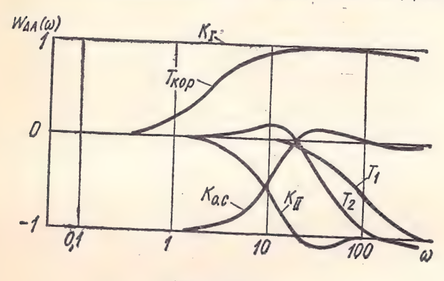


Рис. 9.5

При измерении АЧХ наибольший интерес представляет количество информации об отклонении как отдельных первичных параметров САУ, так и характеристик всей системы в целом.

В случае, если имеет место одновременное случайное отклонение всех первичных параметров [illegible] системы, а также наличие помех измерения, количество информации при измерении можно определить по формуле

$$I_{\Delta Ai} = \frac{1}{2} \log_2 \frac{\sigma^2_{\Delta Ai}(\omega) + \sigma^2_{иA}(\omega)}{\sigma^2_{иA}(\omega)}, \qquad (9.59)$$

где $\sigma_{\Delta Ai} = \omega_{\Delta Ai} \frac{\sigma_{\Delta q_i}}{q_i}$ — среднеквадратическое значение отклонения АЧХ, вызванное отклонением i-го параметра; $\sigma_{иA} = \sqrt{\sigma^2_{Aп} + \sum\limits_{j=1}^{n} \sigma_{\Delta Aj}}$ — среднеквадратическое значение отклонения АЧХ, вызванное шумами; $\sigma_{Aп}$ — среднеквадратическая ошибка прибора измерения АЧХ; $\sum\limits_{j=1;\, j=i}^{n} \sigma_{\Delta Aj}$ — дисперсия АЧХ, обусловленная

отклонениями первичных параметров, не интересующая нас при измерении i-го параметра.

Для сравнения методов контроля по количеству информаций об отклонении отдельных первичных параметров будем полагать, что $\sigma_{An} = 0$, а дисперсии относительных отклонений всех первичных параметров системы равны между собой

$$D = \left(\frac{\sigma_{\Delta k_{\text{I}}}}{k_{\text{I}}}\right)^2 = \left(\frac{\sigma_{\Delta T_{\text{кор}}}}{T_{\text{кор}}}\right)^2 = \left(\frac{\sigma_{\Delta T_1}}{T_1}\right)^2 = \left(\frac{\sigma_{\Delta k_{\text{II}}}}{k_{\text{II}}}\right)^2 = \left(\frac{\sigma_{\Delta T_2}}{T_2}\right)^2 = \left(\frac{\sigma_{\Delta k_{\text{о. с}}}}{k_{\text{о. с}}}\right)^2 .$$

График количества информации, получаемого прибором об отклонениях АЧХ, вызванных отклонением каждого из интересующих нас параметров, приведен на рис. 9.6. Из этого рисунка можно сделать следующие выводы:

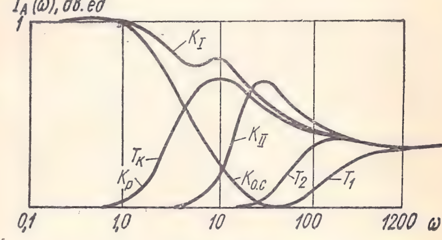


Рис. 9.6

1. Наибольшую информацию измерительный прибор дает об отклонении статических коэффициентов k_{I} и $k_{\text{о. с}}$. Максимум информации получается в области частот

$$\omega = (0 \div 0,1) \frac{1}{\text{c}} .$$

2. Наибольшее количество информации об отклонении коэффициента передачи k_{II} можно получить на частоте

$$\omega \cong 20 \frac{1}{\text{c}} .$$

3. Наибольшее количество информации об отклонении постоянной времени корректирующего контура $T_{\text{кор}}$ можно получить на частоте

$$\omega \cong 10 \frac{1}{\text{c}} .$$

4. Сравнительно меньшее количество информации измерительный прибор дает об отклонении постоянных времени T_1 и T_2.

5. В области частот $\omega = (0,1 \div 100) \frac{1}{\text{c}}$ измерительный прибор дает наибольшее количество информации об отклонениях АЧХ, обусловленной суммарным действием отклонений всех параметров.

Информационная оценка метода и устройства контроля динамической системы, основанного на измерении фазочастотной характеристики (ФЧХ)

Определим веса абсолютных отклонений ФЧХ по параметрам системы, воспользовавшись функциями чувствительности

$$\omega_{\Delta\varphi_{k_1}}(\omega) = Im\,[S_{k_1}(j\omega)] = 0;$$

$$w_{\Delta\varphi T_{\text{кор}}}(\omega) = Im\,[S_{T_{\text{кор}}}(j\omega)];$$

$$w_{\Delta\varphi T_1}(\omega) = Im\,[S_{T_1}(j\omega)];$$

$$w_{\Delta\varphi k_{II}}(\omega) = -\frac{1}{2}\,Im\,[S_{T_k}(j\omega) + S_\xi(j\omega)];$$

$$w_{\Delta\varphi T_2}(\omega) = \frac{1}{2}\,Im\,[S_{T_k}(j\omega) - S_\xi(j\omega)];$$

$$w_{\Delta\varphi k_{\text{о. с}}}(\omega) = -\frac{1}{2}\,Im\,[S_{T_k}(j\omega) + S_\xi(j\omega)]. \qquad (9.60)$$

Как видно из выражения (9.60), веса отклонений ФЧХ по коэффициентам k_{II} и $k_{\text{о. с}}$ равны между собой. Поэтому в дальнейшем изложении вместо двух параметров k_{II} и $k_{\text{о. с}}$ будем рассматривать общий коэффициент второй части системы $k_{II}\cdot k_{\text{о. с}}$ с весом отклонения ФЧХ

$$w_{\Delta\varphi k_{II}k_{\text{о. с}}}(\omega) = \frac{1}{2}\,I_m\,]S_{T_k}(j\omega) + S_\xi(j\omega)]. \qquad (9.61)$$

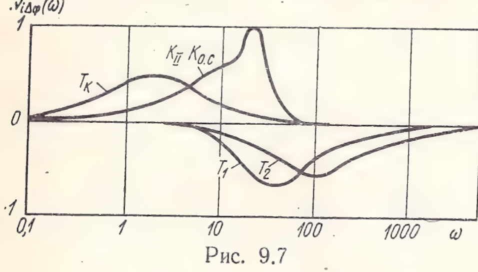


Рис. 9.7

В результате расчета по уравнениям (9.60), (9.61) получим веса отклонений ФЧХ относительно параметров $T_{\text{кор}}$, T_1, k_{II}, $k_{\text{о. с}}$, T_2. Эти зависимости приведены на рис. 9.7.

Как и в предыдущем случае, будем интересоваться количеством информации, получаемой прибором об отклонении данного первичного параметра при измерении ФЧХ с учетом условий одновременного случайного отклонения всех первичных параметров системы

$$I_{\Delta\varphi_i}(\omega) = \frac{1}{2}\log_2\frac{\sigma^2_{\Delta\varphi_i}(\omega) + \sigma^2_{u\varphi}(\omega)}{\sigma^2_{u\varphi}(\omega)}, \qquad (9.62)$$

где $\sigma_{\Delta\varphi_i}(\omega) = \omega_{\Delta\varphi_i}(\omega)_{\Delta q_i}$ — среднеквадратическое значение отклонения ФХЧ, вызванное отклонением i-го параметра; $\sigma_{и\varphi} = \sqrt{\sigma^2_{\varphi n} + \sum\limits_{j=1;\ j\neq i}^{n}\sigma^2_{\Delta\varphi_j}}$ — среднеквадратическое значение отклонения ФЧХ, вызванное шумами; σ_{φ_n} — среднеквадратическая ошибка прибора измерения ФЧХ; $\sum\limits_{j=1;\ j\neq i}^{n}\sigma^2_{\Delta\varphi_j}$ — дисперсия ФЧХ, обусловленная отклонением первичных параметров, не интересующих нас, при измерении i-го параметра.

Графики количества информации об отклонении ФЧХ, вызванном отклонением i-го параметра, получаемой прибором, показаны на рис. 9.8.

Из рассмотрения этого рисунка можно сделать следующие выводы.

1. Количество информации об отклонении параметра на частоте ниже $\omega = 0{,}2\,\frac{1}{c}$ теоретически стремятся к $+\infty$. Это связано с допущением, что ошибка измерительного прибора $\sigma_{\varphi_n} = 0$, а влияние остальных параметров на частотах ниже $\omega = 0{,}2\,\frac{1}{c}$ ничтожно мало. На практике всегда существует ошибка прибора и другие неучтенные помехи, так что количество информации всегда ограничено, что отмечено штриховой линией. Наибольшее количество информации об отклонении $T_{кор}$ лежит в диапазоне частот

$$\omega = (0{,}1 \div 3)\,\frac{1}{c}.$$

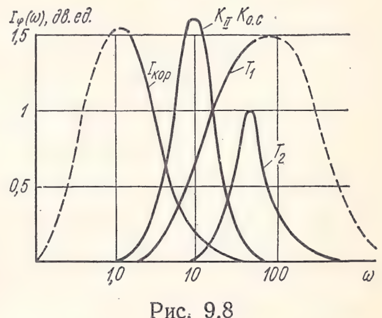


Рис. 9.8

2. Наибольшее количество информации об отклонении параметра $k_{II} \cdot k_{o.c}$ лежит на частоте $\omega \cong 10\,\frac{1}{c}$.

3. Количество информации об отклонении параметров также ограничено из-за несправедливости условия $\sigma_{\varphi_n} = 0$.

4. Графики количества информации для различных параметров имеют максимумы, разнесенные на оси частот. Диапазоны частот, в которых сосредоточена информация об отклонениях отдельных параметров при измерении ФЧХ, разделяются на оси частот гораздо легче, чем при измерении отклонений АЧХ от номинальной.

Информационная оценка метода и устройства контроля динамической системы, основанного на измерении переходной функции системы

Веса отклонений переходной функции по рассмотренным ранее шести первичным параметрам можно определить путем моделирования на электронной модели непрерывного действия (рис. 99).

При подаче на вход модели единичного скачка напряжения на выходе модели системы получим переходную функцию, а на выходе преобразованных цепей — веса отклонений переходной функции по каждому из шести параметров.

На рис. 9.10 приведены графики весов отклонений переходной функции, полученные путем моделирования на вычислительной машине непрерывного действия МН-7.

Количество информации об отклонении данного параметра, получаемой прибором измерения отклонений переходной функции при условии, что имеет место одновременное случайное отклонение всех параметров системы,

$$I_{\Delta h_i} = \frac{1}{2}\log_2 \frac{\sigma^2_{\Delta h_i}(t) + \sigma^2_{иh}(t)}{\sigma^2_{иh}(t)},$$

где $\sigma_{\Delta h_i}(t) = w_{\Delta h_i}(t)\dfrac{\Delta q_i}{q_i}$ — среднеквадратическое значение отклонения переходной функции, вызванное отклонением i-го параметра системы; $\sigma_{иh}(t) = \sqrt{\sigma^2_{h\text{п}} + \sum\limits_{j=1;\, j\neq i}^{n} \sigma^2_{\Delta h_j}}$ — среднеквадратическое значение отклонения переходной функции, вызванное шумами; $\sigma_{h\text{п}}$ — среднеквадратическая ошибка прибора измерения переходной функции; $\sum\limits_{j=1;\, j\neq i}^{n} \sigma^2_{\Delta h_j}$ — дисперсия переходной функции, обусловленная отклонениями первичных параметров, не интересующая нас при измерении отклонения переходной функции, обусловленное отклонением i-го параметра.

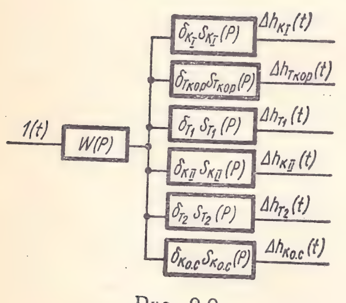


Рис. 9.9

На рис. 9.11 показаны графики количества информации, получаемой прибором измерения отклонения переходной функции и вызванного отклонением интересующего нас первичного параметра.

Из рассмотрения этого рисунка можно сделать следующие выводы.

1. Наибольшее количество информации об отклонении коэффициента усиления k_I можно получить по истечении $0{,}5 \div 0{,}6$ с. от начала переходного процесса.

2. Максимум информации об отклонении постоянной времени $T_{кор}$ лежит в диапазоне времени 0—0,2 с.

3. Максимальное количество информации об отклонении параметра k_{II} лежит в интервале времени 0—0,1 с.

4. Наибольшее количество информации об отклонении коэффициента $k_{о.с}$ лежит в диапазоне 0,15—0,5 с.

5. Количество информации об отклонении параметров высокочастотных звеньев T_1, T_2 очень мало и сосредоточено в начале переходного процесса.

6. Если измеряются значения переходной функции в отдельные моменты времени, то, как это видно из рисунка, целесообразно выбрать такие точки, в которых отклонения переходной функции несут наибольшее количество информации об отклонении параметров.

Сравнительная информационная оценка методов контроля.

Для сравнительной информационной оценки методов и устройств контроля динамической системы по АЧХ, ФЧХ и переходной функции воспользуемся графиками количества информации, получаемой при измерении характеристик системы.

Из сравнения этих графиков можно сделать следующие выводы:

1. Наибольшее количество информации об отклонении всех параметров системы дает устройство, измеряющее ФЧХ; наименьшее — устройство, измеряющее переходную функцию, ФЧХ не дает информации об отклонении коэффициента k_I, а АЧХ и переходная

функция дают одинаковое количество информации об отклонении этого коэффициента.

2. Графики количества информации по k_I имеют, кроме указанных в таблице максимальных значений, экстремум (максимум). При этом экстремальное значение $I_{\Delta A}(\omega)$ на частоте $10\frac{1}{c}$ соответствует экстремальному значению $I_{\Delta h}(\omega)$ в момент времени 0,05 с, а максимальное значение количества информации $I_{\Delta A}(\omega)$, начиная

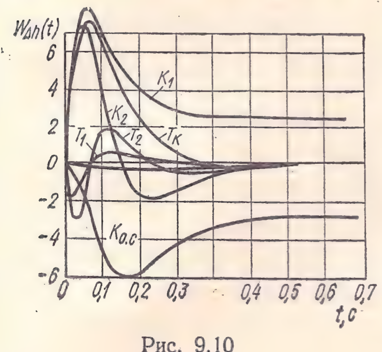


Рис. 9.10

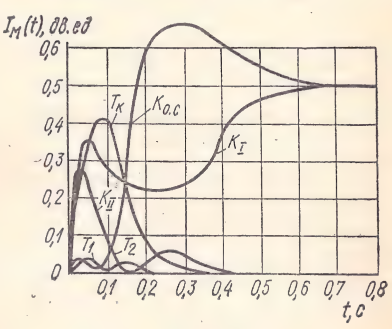


Рис. 9.11

с частоты $\omega = 0{,}71\frac{1}{c}$ и до $\omega = 0$, соответствует максимальному количеству информации $I_{\Delta h}(t)$ в момент времени $t > 0{,}6$с.

3. Графики количества информации об отклонении АЧХ, ФЧХ и переходной функции имеют явно выраженные максимумы. При этом максимум $I_{\Delta A}(\omega)$ при $\omega = 10\frac{1}{c}$ соответствует максимуму $I_{\Delta\varphi}(\omega)$ на $\omega = 1\frac{1}{c}$ и максимуму $I_{\Delta h}(t)$ в момент времени 0,1 с.

4. Графики количества информации об отклонении коэффициента k_{II} также имеют явно выраженные максимумы. Максимум $I_{\Delta A}(\omega)$ на $\omega = 30\frac{1}{c}$ соответствует максимуму $I_{\Delta\varphi}(\omega)$ на $\omega = 10\frac{1}{c}$ и максимуму $I_{\Delta h}(t)$ в момент времени 0,02с.

5. Графики количества информации $I_{\Delta A}(\omega)$ об отклонении имеют максимальное значение в диапазоне $\omega\,(0 \div 0{,}7)\ \frac{1}{c}$, а $I_{\Delta h\,(t)}$ имеют максимальное значение в момент времени 0,35с.

6. Количество информации об отклонениях параметров T_1 и T_2 при измерении АЧХ и ФЧХ в области высоких частот получается малым, не имеющим практического значения; при измерении переходной функции количество информации об отклонении параметров T_1 и T_2 получается сравнительно малым и расположено в начале переходного процесса.

7. Отклонения параметров $T_{кор}$, k_{II}, $k_{о.с}$, T_1, T_2 легче различать при контроле ФЧХ, а отклонения параметров k_I, $k_{о.с}$ — при контроле АЧХ и переходной функции.

СПИСОК ЛИТЕРАТУРЫ

1. Брежнев Л. И. Отчетный доклад ЦК КПСС XXV съезду КПСС. М., Госполитиздат, 1976.

2. Элементы технической кибернетики. Терминология. М., «Наука», 1968.

3. Хартли Р. Л. Передача информации. — В сб.: Теория информации и ее приложения. М., Физматгиз, 1959.

4. Котельников В. А. О пропускной способности эфира и проволоки в радиосвязи. Изд-во МГУ, 1933.

5. Котельников В. А. Теория потенциальной помехоустойчивости. М., Госэнергоиздат, 1946.

6. Буга Н. Н. Основы теории связи и передачи данных, ч. I. Л., 1968.

7. Харкевич А. А. Спектры и анализ. М., Госэнергоиздат, 1963.

8. Тихонов В. И. Статистическая радиотехника. М., «Советское радио», 1966.

9. Солодов А. В. Теория информации и ее применение к задачам автоматического управления и контроля. М., «Наука», 1967.

10. Фельдбаум А. А. и др. Теоретические основы управления. М., Физматгиз, 1963.

11. Ицхоки Я. С. Импульсные устройства. М., «Советское радио», 1959.

12. Кедрус В. А. Основы телемеханики, ч. I. Харьков, Изд-во ХВКИУ, 1969.

13. Вентцель Е. С. Теория вероятностей. М., Физматгиз, 1958.

14. Левин Б. Р. Теория случайных процессов и ее применение в радиотехнике. М., «Советское радио», 1960.

15. Терентьев С. Н. Основы общей теории связи. Харьков, Изд-во ХВКИУ, 1968.

16. Кузьмин И. В. Теоретические основы информационной техники. Харьков, Изд-во ХВКИУ, 1969.

17. Назаров М. В., Кувшинов Б. И., Попов О. В. Теория передачи сигналов. М., «Связь», 1970.

18. Гмурман В. Е. Введение в теорию вероятностей и математическую статистику. М., «Высшая школа», 1963.

19. Железнов Н. А. Некоторые вопросы теории информационных электрических систем. Л., Изд-во ЛКВВИА им. А. Ф. Можайского, 1960.

20. Шастова Г. А. Кодирование и помехоустойчивость передачи телемеханической информации, М., «Энергия», 1966.

21. Математика, ее содержание, методы и значение. М., Труды АН СССР, 1956.

22. Пугачев В. С. Теория случайных функций. М., Физматгиз, 1962.

23. Шеннон К. Работы по теории информации и кибернетике. М., «Иностранная литература», 1963.

24. Фано Р. Передача информации. Статистическая теория связи. М., «Мир», 1965.

25. Голдман С. Теория информации. М., «Иностранная литература», 1957.

26. Клюев Е. В. Информационные основы передачи сообщений. М., «Советское радио», 1966.

27. Зеленый Д. М. Основы теории информации и оптимального приема. Л., Изд-во ВМА, 1966.
28. Темников Ф. Е., Афонин В. А, Дмитриев В. И. Теоретические основы информационной техники. М., «Энергия», 1971.
29. Карнап Р. Значение и необходимость. М., «Иностранная литература», 1959.
30. Харкевич А. А. О ценности информации.— «Проблемы кибернетики», вып. 4. М., Физматгиз, 1960.
31. Хинчин А. Я. Понятие энтропии в теории информации. Успехи математ. наук, вып. 3. М., Физматгиз, 1953.
32. Беляков В. В., Расщепляев Ю. С., Соколов В. А. Основы теории прохождения сигналов. Ростов, Изд-во РВКИУ, 1970.
33. Латхи Б. П. Системы передачи информации. М., «Связь», 1971.
34. Оливер Б. Эффективное кодирование. — В сб.: Теория информации и ее приложения. М., Физматгиз, 1959.
35. Харкевич А. А. Очерки общей теории связи. М., Гостехиздат, 1955.
36. Васильев П. Р., Шастова Г. А. Передача телемеханической информации. М., «Энергия», 1960.
37. Харкевич А. А. Борьба с помехами. М., Физматгиз, 1965.
38. Катков Ф. А., Попов А. Б. Частотные системы телеуправления по занятым каналам связи. Киев, Гостехиздат, 1963.
39. Катков Ф. А. Многочастотные узкополосные системы телеуправления. Киев, Гостехиздат, 1960.
40. Котов П. А. Повышение достоверности передачи цифровой информации. М., «Связь», 1966.
41. Боуз Р. С., Рой-Чоудхури Д. К. Об одном классе двойных групповых кодов с исправлением ошибок.— «Кибернетический сборник», вып. 6. М., «Иностранная литература», 1963.
42. Удалов А. П., Супрун Б. А. Избыточное кодирование при передаче информации двоичными кодами. М., «Связь», 1964.
43. Питерсон У. Коды, исправляющие ошибки. М., «Мир», 1964.
44. Мартынов Ю. М. Обработка информации в системах передачи данных. М., «Связь», 1969.
45. Каневский З. М. Передача сообщений с информационной обратной связью. М., «Связь», 1969.
46. Каневский З. М., Финкельштейн М. И. Флюктуационная помеха и обнаружение импульсных радиосигналов. М., Госэнергоиздат, 1963.
47. Левин Б. Р. Теоретические основы статистической радиотехники. М., «Советское радио», 1968.
48. Гуткин Л. С. Теория оптимальных методов радиоприема при флюктуационных помехах. М., Госэнергоиздат, 1961.
49. Касаткин А. С., Кузьмин И. В. Оценка эффективности автоматизированных систем контроля. М., «Энергия», 1967.
50. Красовский А. А., Поспелов Г. С. Основы автоматики и технической кибернетики. М., Госэнергоиздат, 1962.
51. Новицкий П. В. Основы информационной теории измерительных устройств. М., «Энергия», 1968.
52. Кузьмин И. В., Кедрус В. А. Критерии оценки качества функционирования и пути оптимизации систем автоматизированной проверки измерительных средств.— В сб.: Автоматизированные системы управления и приборы автоматики, вып. 26. Харьков, Изд-во ХГУ, 1972.
53. Синдеев И. М. О выборе параметров, определяющих состояние технического устройства при автоматическом контроле. Труды ВВИА им. проф. Н. Е. Жуковского, вып. 1020. М., 1963.
54. Кузьмин И. В. Оценка эффективности автоматических систем контроля и управления. Харьков, Изд-во ХВКИУ, 1966.
55. Кузьмин И. В. Элементы вероятностных моделей АСУ. М., «Советское радио», 1975.
56. Кузьмин И. В. Оценка эффективности и оптимизация АСКУ. М., «Советское радио», 1971.

57. Синдеев И. М. Некоторые общие методы автоматического контроля состояния динамических систем и поиска неисправностей. Труды ВВИА им. проф. Н. Е. Жуковского. М., 1962.

58. Быковский М. Л. Основы динамической точности электрических и механических цепей. М., Изд-во АН СССР, 1958.

59. Чан С. С. Л. Информационный критерий для замкнутых систем автоматического регулирования. Труды ИФАК. М., Изд-во АН СССР, 1961.

60. Колмогоров А. Н. Теория передачи информации. Сессия АН СССР по научным проблемам автоматизации производства. Пленарные заседания. М., Изд-во АН СССР, 1957.

61. Финк Л. М. Теория передачи дискретных сообщений. М., «Советское радио», 1963.

62. Кузьмин И. В. Проектирование автоматизированных телемеханических систем контроля и управления. Ч. III. Оператор в системе контроля и управления. Харьков, Изд-во ХВКИУ, 1970.

63. Орнатский П. П. Автоматические измерения и приборы. Киев, «Вища школа», 1973.

64. Зюко А. Г., Коробов Ю. Ф. Теория передачи сигналов. М., «Связь», 1972.

65. Философский словарь, ред. Розенталь М. М. и Юдина П. Ф. М., Госполитиздат, 1963.

66. Гуров В. С. и др. Передача дискретной информации и телеграфия, М., «Связь», 1974.

СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ

1. Кульбак С. Теория информации и статистика. М., «Наука», 1967.

2. Гойхман Э. Ш., Лосев Ю. И. Передача информации в АСУ. М., «Связь», 1971.

3. Дружинин Н. К. Логика оценки статистических гипотез. М., «Статистика», 1973.

4. Ицхоки Я. С., Овчинников Н. И. Импульсные и цифровые устройства. М., «Советское радио», 1972.

5. Шестов Н. С. Выделение оптических сигналов на фоне случайных помех. М., «Советское радио», 1967.

6. Мищенко В. А. Метод селектирующих функций в нелинейных задачах контроля и управления. М., «Советское радио», 1973.

7. Данилин Н. С., Бакланов О. Д., Загоровский Ю. И. Теория и методы неразрушающего контроля радиоэлектронных схем. М., Изд-во Министерства обороны СССР, 1974.

8. Глушков В. М. Введение в АСУ. Киев, «Технiка», 1972.

9. Захаров В. Н., Поспелов Д. А., Хозацкий В. Е. Системы управления. М., «Энергия», 1972.

10. Пономарев В. М. и др. Основы автоматического регулирования и управления. М., «Высшая школа», 1974.

11. Глушков В. М. и др. Энциклопедия кибернетики, т. 1, 2. Киев, УСЭ, 1974.

12. Галлагер Р. Теория информации и надежная связь. М., «Советское радио», 1974.

13. Бовбель Е. И., Данейко И. К., Изох В. В. Элементы теории информации. Минск, Изд-во БГУ им. В. И. Ленина, 1974.

14. Рози А. М. Теория информации и связи. М., «Энергия», 1971.

15. Стратонович Р. Л. Теория информации. М., «Советское радио», 1975.

16. Полетаев И. А. Сигнал. М., «Советское радио», 1958.

17. Николаев В. И. Информационная теория контроля и управления. Л., «Судостроение», 1973.

18. Четвериков В. Н. Преобразование и передача информации в АСУ. М., «Высшая школа», 1974.

ОГЛАВЛЕНИЕ

Предисловие 3
Введение 5

Глава I. **Информация и информационные системы** 8

1.1. Основные понятия и определения 8
1.2. Классификация информационных систем 12
1.3. Предмет и метод теории информации 13
Контрольные вопросы 15

Глава II. **Математические модели детерминированных сигналов** 16

2.1. Общая характеристика сигналов 16
2.2. Частотное представление детерминированных сигналов 21
1. Периодические сигналы 21
2. Непериодические сигналы 31
3. Сопоставление спектров периодических сигналов и соответствующих непериодических сигналов 32
4. Энергетическое толкование спектра сигнала 33
5. Практическая ширина спектра сигнала 36
2.3. Переход от преобразования Фурье к преобразованию Лапласа 41
2.4. Примеры 42
Контрольные вопросы 48

Глава III. **Математические модели случайных сигналов** 48

3.1. Случайные сигналы и их вероятностные характеристики 48
3.2. Числовые характеристики случайного процесса 51
3.3. Стационарные случайные процессы 54
3.4. Свойства корреляционной функции стационарного случайного процесса 56
3.5. Эргодичность стационарных процессов 59
3.6. Спектральная плотность стационарного случайного процесса 60
3.7. Эффективная ширина спектра случайного процесса 66
3.8. Кодирование сигналов. Модели сигналов при кодоимпульсной модуляции 66
3.9. Квантование по времени 71
3.10. Квантование по уровню 78
3.11. Примеры 83
Контрольные вопросы 97

Глава IV. **Информационные модели сигналов** 98

4.1. Количество информации и неопределенность. Энтропия как мера неопределенности 98
4.2. Свойства энтропии дискретных сообщений 101
4.3. Энтропия непрерывных сообщений 103

4.4. Энтропия сложных сообщений 105
4.5. Количество информации при неполной достоверности сообщений . . . 108
1. Дискретные сообщения 108
2. Непрерывные сообщения 111
4.6. Энтропия и количество информации при статистической зависимости элементов сообщений 114
4.7. Избыточность сообщений 116
4.8. Преимущества и недостатки статистической меры количества информации 118
4.9. Примеры 119
Контрольные вопросы 130

Глава V. **Передача информации** 131

5.1. Обобщенные характеристики сигналов и информационных каналов . . 131
5.2. Скорость передачи информации и пропускная способность дискретного канала без помех 134
5.3. Скорость передачи информации и пропускная способность дискретного канала с помехами 136
5.4. Скорость передачи информации и пропускная способность непрерывного канала с помехами 141
5.5. Примеры 147
Контрольные вопросы 154

Глава VI. **Эффективность информационных систем** 154

6.1. Критерии оценки эффективности информационных систем 154
6.2. Способы повышения эффективности информационных систем 155
6.3. Перераспределение плотностей вероятностей элементов сообщения . . 156
6.4. Декорреляция сообщений 158
6.5. Оптимальное статистическое кодирование 160
6.6. Примеры 164
Контрольные вопросы 167

Глава VII. **Помехоустойчивость информационных систем** 168

7.1. Общая характеристика помех в системах передачи информации 168
7.2. Критерии оценки помехоустойчивости информационных систем 170
7.3. Способы повышения помехоустойчивости информационных систем . . 174
7.4. Помехоустойчивость различных видов модуляции 176
7.5. Помехоустойчивое кодирование. Основные принципы помехоустойчивого кодирования 181
7.6. Связь исправляющей способности кода с кодовым расстоянием 184
7.7. Построение кодов с заданной исправляющей способностью 186
7.8. Геометрическая модель комбинаций двоичного кода 188
7.9. Показатели качества корректирующих кодов 189
7.10. Классификация помехоустойчивых кодов 189
7.11. Код с четным числом единиц 191
7.12. Код с удвоением элементов и инверсный код 191
7.13. Коды Хемминга 192
7.14. Циклические коды 194
7.15. Системы с обратной связью 198
7.16. Примеры 200
Контрольные вопросы 208

Глава VIII. **Элементы теории оптимального приема и статистических решений** 209

8.1. Частотная фильтрация 209
8.2. Метод накопления 211
8.3. Корреляционный метод 213

8.4. Согласованная фильтрация 215
8.5. Сущность основной задачи приема сигналов при наличии помех . . . 216
8.6. Обнаружение сигнала 219
8.7. Различение сигналов 226
8.8. Синтез структуры решающего устройства 228
8.9. Восстановление сигналов 230
8.10. Примеры 234
Контрольные вопросы 245

Глава IX. **Информационная оценка автоматизированных систем контроля и управления** 246

9.1. Общая характеристика автоматизированных систем контроля и управления (АСКУ) 246
9.2. Оценка степени неопределенности состояния объекта контроля . . . 247
9.3. Информационная оценка точности результата контроля 250
9.4. Информационная способность устройства контроля и информационный к. п. д. процесса контроля 252
9.5. Пропускная способность процесса контроля 254
9.6. Энтропия и информация в системах автоматического управления . . . 255
9.7 Оценка эффективности АСКУ 259
9.8. Информационное обоснование оптимального алгоритма контроля и поиска неисправностей 261
9.9. Оценка методов и устройств контроля динамической системы стабилизации летательного аппарата 265

Список литературы 274
Список дополнительной литературы 276

Иван Васильевич Кузьмин
Валентин Александрович Кедрус

Основы теории информации и кодирования

Киев
Головное издательство
издательского объединения «Вища школа»

Редактор Ж. Г. Давиденко
Обложка художника Б. Н. Жеребцова
Художественный редактор С. П. Духленко
Технический редактор Л. Ф. Волкова
Корректор О. А. Савицкая

Информ. бланк № 2464.

Сдано в набор 23.08.76 г. Подписано в печать 14.01.77 г. Формат 60×90[1]/16. Бумага типографская № 3. 17,5 печ. л. 16,95 уч.-изд. л. Тираж 13000 экз. Изд. № 2323. БФ 08169. Зак. № 6-34. Цена 85 коп.

Головное издательство издательского объединения «Вища школа», 252054, Киев-54, Гоголевская, 7.

Книжная фабрика им. М. В. Фрунзе Республиканского производственного объединения «Полиграфкнига» Госкомиздата УССР, Харьков, Донец-Захаржевская, 6/8.

85 коп.

И. Б. КУЗЬМИН, В. А. КЕДРУС